

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:43:59 ; Search time 42 seconds  
(without alignments)  
826.593 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886  
Sequence: 1 MNKILFIPLFFSSVLETFEA.....EAGQYMGNIWTFPSSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1558.5	82.6	360	2 D56617	cfab protein precu
2	949.5	50.3	363	2 S49539	Cold protein precu
3	930.5	49.3	364	2 S57937	probable fimbrial
4	275.5	14.6	359	2 AE0541	probable fimbrial
5	120	6.4	1335	2 G90975	probable factor (1
6	120	6.4	2660	2 E85822	probable invas
7	114	6.0	650	2 S22835	alpha-agglutinin
8	109	5.8	2383	2 D64962	probable membrane
9	108	5.7	524	2 S55097	probable membrane
10	108	5.7	2529	2 B64635	toxin-like outer m
11	107	5.7	304	2 A64904	probable fimbrial
12	107	5.7	304	2 C90892	probable adhesin
13	107	5.7	304	2 E85725	probable adhesin
14	106	5.6	671	2 A38109	autolysin - Entero
15	105.5	5.6	843	2 AC2507	hypothetical prote
16	105	5.6	598	2 AB1236	hypothetical prote
17	104.5	5.5	1441	2 B86807	hypothetical prote
18	103.5	5.5	691	2 B75622	hypothetical prote
19	103.5	5.5	1269	2 A90267	proteinase related
20	103.5	5.5	2399	2 H71879	toxin-like outer m
21	102	5.4	692	2 G90284	hypothetical prote
22	101	5.4	1238	2 AH0038	probable exported
23	100	5.3	1752	2 T48965	hypothetical prote
24	99.5	5.2	5627	2 C83339	hypothetical prote
25	99	5.2	802	2 AH1580	phenylalanyl-tRNA
26	99	5.2	856	2 B81399	probable periplasm
27	99	5.2	1849	2 C41859	Iga-specific metal
28	98.5	5.2	608	2 H64473	hypothetical prote
29	98.5	5.2	4919	2 T31105	hypothetical prote

30	98	5.2	587	2 AC1510	internalin protein
31	98	5.2	758	2 S47866	SPR21 protein - ye
32	98	5.2	1014	2 S37405	cytotoxic necrotiz
33	98	5.2	1157	2 S38160	NUP133 protein - y
34	97.5	5.2	1797	2 F69195	cell surface glyco
35	97	5.1	454	2 T01337	hypothetical prote
36	96.5	5.1	2364	2 I40884	cytotoxin L - Clos
37	96	5.1	992	2 T28421	probable DNA-direc
38	95.5	5.1	713	1 A185XR	cyclomaldextrin
39	95.5	5.1	895	2 T02597	Mutator-like trans
40	95	5.0	413	1 S28602	translation releas
41	95	5.0	561	2 AB4113	transposase (12) B
42	95	5.0	2334	2 S32920	cell wall-associat
43	95	5.0	5188	2 B85547	probable RTX fam1
44	94.5	5.0	835	2 E71691	outer membrane ass
45	94.5	5.0	858	2 T47223	replication licens

## ALIGNMENTS

## RESULT 1

D56617 cfab protein precursor - Escherichia coli plasmid NTP113

C:Species: Escherichia coli

C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 10-Dec-1999

C:Accession: D56617 R:Jordl, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastera, W.

DNA Seq. 2, 257-263, 1992

A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of

A:Reference number: A56617; MUID:92329981; PMID:1352712

A:Accession: D56617

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <OR>

A:Cross-references: GB:M55661; NID:q145507; PIDN:AAC41417.1; PID:q145511

A:Experimental source: enterotoxigenic strain, CFA/I-ST plasmid NTP113

A>Note: sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P:108972)

C:Genetics:

A:Gene: cfab

A:Genome: cfab

C:Superfamily: Escherichia colonizing factor antigen cfab

Query Match 82.6%; Score 1558.5; DB 2: Length 360;

Best Local Similarity 81.2%; Pred. No. 1.4e-109;

Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;

QY	1	MNKLFIPLFFSSVLETFEAVSADKIPGDEITNIFGPRDRNSSPKHNLNNHITAYSE	60
DB	1	MNKLFIPLFFSSGFTFAVSADKNPGEENMTNIGPHDRGSSPTNINLSYLAVNG	60
QY	61	SHLDRMTFLCLSSNITNGACPTSENSSSSSEETNITQTFEKRSLIRELOIKY	120
DB	61	SHLDRMTFLCLSSNITNGACPTSENSSSSSEETNITQTFEKRSLIRELOIKY	120
QY	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATLARK	180
DB	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATLARK	180
QY	181	RYSEETGYTNTITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCY	240
DB	181	RYSEETGYTNTITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCY	240
QY	241	DGYSTNSSLETRFODNNKSGKRYLRIINDTKEIATLSLLAGKSLPTNGSLNI	300
DB	241	DGYSTNSSLETRFODNNKSGKRYLRIINDTKEIATLSLLAGKSLPTNGSLNI	300
QY	301	ADAASLETWNRITAVTMEISVPVLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
DB	301	ADAASLETWNRITAVTMEISVPVLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
QY	361	L 361	

```

Db          360 L3360

RESULT 2
C:Cood protein precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
C:Accession: S49539
R:Roehlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.
Mol. Microbiol. 12, 387-401, 1994
A:Title: Cooc and Coocd are required for assembly of CS1 pili.
A:Reference number: S49538; MUID:94344028; PMID:7915003
A:Accession: S49539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-363 <PRO>
A:Cross-references: EMBL:X76908; NID:9488735; PIDN:CAA54230.1; PID:9488737
Superfamily: Escherichia colonizing factor antigen cfae

Query Match          50.3%; Score 949.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 7.7e-64;
Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

QY 3 KILFIPLFFSVVFTFAVSADKIPGDE-SITNIF-GPR-DRNESPKHILNNHITAY 58
Db 2 KKITFL-SSIFSAVSAGRYPETVGNLTSFAPRIIDRSVQSEIYNIFTNHVAGY 57
QY 59 SESEITLIDRMFLCLSSHNTLNGACPTSENSSSSVS-GETNITLQTEFRSLIKRELQI 117
Db 58 SLSHSLVDRIVFLCTSSSNPNVNGACPTI--GTSGVQVGTITITLQTEFRSLIKRNTNL 114
QY 118 KGYIQLLFKSYNCSG--LITNSAHFNCKNKA-ASGASLYIYPAGELKNIPEGCIWDAT 174
Db 115 AGNKKPRTWENSCDFSNLMVNLNSKSWSCGANGELTILNLYIPAGEIINKLPFGIEWAT 174
QY 175 LKLEVKKRRYSETP---YGTATINTITIKLTDKNGIQLIOPKSPDARDVNLNRPFGGGYI 230
Db 175 LILKLS-RYGVSSSTHNGNTVNTITVDLTDKNGIQWLPGHSPRPDLNLRPIGNKYIS 233
QY 231 GRNSVDMECFYDGYSTNSSLSLEIRFQDNNPKSGDKFYLRKINDTKEIATYLLSLLAGKSL 290
Db 234 GSNSLDMCFYDGYSTNSDSVYIKFQDDNPNTSSEVNLTKIG-GTEKLPYAVSLMGEKIF 292
QY 291 TPPTGTSINIDASLEENMRITAVMPETLSVYVLCPGQLQIDAKYENEAQYMGNI 350
Db 293 YPVNGQSTTINDSSVLETNMRITAVAPPEVNVPLCPARILLNADVNADAGQYSQI 352
QY 351 NVTFPPSSQTL 361
Db 353 YTFETPSEVNL 363

RESULT 3
C:Coed protein precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999
C:Accession: S57937
R:Roehlich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.
submitted to the EMBL Data Library, January 1995
A:Description: The genes for CS2 pili of enterotoxigenic Escherichia coli and their inter-
A:Accession: S57937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <PRO>
A:Cross-references: EMBL:Z47800; NID:9897725; PIDN:CAA87763.1; PID:9897729
C:Genetics:
A:Start codon: TTG
C:Superfamily: Escherichia colonizing factor antigen cfae

Query Match          49.3%; Score 930.5; DB 2; Length 364;

```

Best Local Similarity 48.4%, Pred. No. 2.1e-62;  
Matches 177, Conservative 64, Mismatches 118, Indels 7, Gaps 3;

```
OY      1 MNKLIFLTLEFSSVLTFPAVSADKIFDESLINIEGPPDRN--ESSPKHNLINNHITAY 58
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 MKKVIYFLSMFLCQVYQGSWHTVVEGSLINKTESIPIDRSAAASYPAHYIFEHVAGY 60
OY      59 SESELTIDRMFLCSLHNTLNGACPISSENPSSSSVSGENITIQFTEKRSILKREIQIK 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61 NKHSLIDRMFLCWSLSDASKGACPIGSEKSS--GGEINIKILFEKRSILKARTLNK 118
OY      119 GYKOLLERSVNC--PSGLTILSAHFNCNKAASGASLYIYPAGELKNLPFGGIWADTL 175
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 GYKRFLEESDPCIHVYVKMNLNHSHTVACVSGFTFGVPEFTLYIPQGEIDGLTGIEWEATL 178
OY      176 KLRKRRRSEYGYTITNITIKLFDKGNIOIWLPOFQSDARVDNLNRPFGGTYIGNSV 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      179 ELARKHYDHWGKIKYKNTIYDLDTKGNIOIWLPIKFSIDPRIDLNLPBEGNGKTSQSNVL 238
OY      236 DMCFYDYGTINSSLEIRFODNNPKSDGKEYLRKINDTKEIATVLSLLAGRSLEPTNG 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      239 EMCYTDGYSTHSGSIEKRFODDSQGTNNENYLIKTGEPILKPKLYLSLLGREFYPNNG 298
OY      296 TSLNIADAASLETWNNTATVMPLEISVPYLCWGRQLQDAKYNENPAQGYMGININTFT 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      299 EAPITNTSSLEFIWMNRKISVLPQISIPVLCWPNALPENSELNPPAGEYSGILNVTFT 358
OY      356 PSSQTL 361
      : : : : :
Db      359 PSSSSL 364
```

RESULT 4  
AE0541  
probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serovar  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AE0541  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerthon, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AE0541  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-359 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:g16501589; GSPDB:GN00176  
C:Genetics:  
A:Gene: tcfD

Query Match 14.6%; Score 275.5; DB 2; Length 359;  
Best Local Similarity 29.7%; Pred. No. 3.2e-13;  
Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

```
OY      70 FLCLSSHNTLNGACPISSENPSSSVSGENITIQFTEKRSILKREIQIKGYKOLLERSVN 129
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      79 WVCSSNNENEGACEETHLWVYAFGAYSKIRLFRFQISHAETTL-----ILLGSVR 131
OY      130 --CGSGTLNSAHNCKKNMAAS---GSLTXLYIPAGELKNLPFGGIWADTLKL-RVKKRY 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      132 DACTGV-----INNNAAACOWGSRSLKRLRIPSEELAKIPISGIMKATLVLDYIOWMG 183
OY      184 SETYGTATITITIKL--KGNLIQWLPQC-KSDARVDNLNRPFGGTYIGRSVDMCFY 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      184 DDLPGTSTDTITLVNVDHFAENALYPRQGTATPRVDNLNHNMAASQMSGGRANLDMCLY 243
OY      241 DGVSTNSSLEIRFODNNPKSDGKEYLRKINDTKEIATVLSLLAGKSLTPNGISLNT 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      244 DG-GVKRRLSRIKIEGNSKSGTGFQVYIKSSADT--IDVAVSNNGYGRSLPVRIGVEEFL 300
```

OY 301 ADAASLETNNRITAVTMEISVPLVLCMPGRLOLDAK---VENPAGQVMGNINVTTPS 357  
 Db 301 DNVDKRAATR-----PVVLEQROARVCPVPLTTLTTPFNIRKRSRSGEYGTIVTMKMG 355  
 OY 358 SQ7 360  
 Db 356 TOT 358

## RESULT 5

G90975

Probable factor [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)  
 C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C/Accession: G90975

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
 A/Reference number: A99629; MUID:21156231; PMID:11258796

Accession: G90975

Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-1335 &lt;HAY&gt;

A/Cross-references: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RMD 0509952

C/Genetics:

A:Gene: ECs2775

Query Match 6.4%; Score 120; DB 2; Length 1335;  
 Best Local Similarity 20.4%; Pred. No. 0.86;  
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

OY 13 SSVLFFPANSADKIPGDESTITNIFGPRDRNESSPKHNILNNHTAY---SESHLTIDRMT 69  
 Db 746 SAKIATLSASNNQVLANENAAVTSVNADEGS---NPINDHTVTFVAVLSGATSFNQNM 802  
 OY 70 FLCLSSHNTLNGACPTSENPSSSSVSGETNITLOTFEKRSLI-----K 112  
 Db 803 ---TAKTDVNGLA-TFDLKSSKQEDNTVETILENGVKQTLIVSPGDSSTAQVDLOKSK 857  
 OY 113 RELQIKGY-KOLLEKSVNCPGSLTTLNSAHFNCKRNAASGASLYIYPAGELNLPFGGIW 171  
 Db 858 NEVVADGNDGATMTATVBDKAGNLDVKTFFVNSAAKLSQTEVNSHD-----GIA 910  
 OY 172 DATLKLRVRRRYSETGTYTINTIKLTDKGNIQIWLPOKRSARVDLNRPTGGGYIG 231  
 Db 911 TATLT-----SLKNGDYVTVASVSSGQANQOVIFIGDQSTALTLVS-PSGDITV-- 960  
 OY 232 RNSVDMCFYDGYSTNSSLE---IRFODNPNKSDCKEFLRKINDDTKEIAYTLISLLAG 287  
 Db 961 -----TNTAPLHMTATLQDKNPNPKDKKEITFSVND-----VASFFSISNSG 1003  
 OY 288 KSLTPINGTSL-----NIADAASLETNNRITAVTMEISVPLC 327  
 Db 1004 KGMTDSNGTAIASLTGLAGTHMITARLANSVSDQPMTFVADKRAVVAVLQTSKAEII 1063  
 OY 328 WPG--RLQDRAKVENPEAGQYMGINIVTTPS 357  
 Db 1064 GNGVDETTTLTATVKDP-FDNVKNLSVVERFS 1094

## RESULT 6

E85822

Probable Invasin Z3135 [imported] - Escherichia coli (strain O157:H7, substrain EDU933)  
 C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: E85822

C/Accession: E85822

R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Iller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoussis, K.; Apodaca,  
 Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: E85822  
 A/Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2660 <STO>  
 A/Cross-references: GB:AE005174; NID:g12516151; PIDN:AA057041.1; GSPDB:GN00145; UWGP:  
 A/Experimental source: strain O157:H7, substrain EDU933  
 C/Genetics:  
 A:Gene: Z3135

Query Match 6.4%; Score 120; DB 2; Length 2660;  
 Best Local Similarity 20.4%; Pred. No. 2.1;  
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

OY 13 SSVLFFPANSADKIPGDESTITNIFGPRDRNESSPKHNILNNHTAY---SESHLTIDRMT 69  
 Db 737 SAKIATLSASNNQVLANENAAVTSVNADEGS---NPINDHTVTFVAVLSGATSFNQNM 793  
 OY 70 FLCLSSHNTLNGACPTSENPSSSSVSGETNITLOTFEKRSLI-----K 112  
 Db 794 ---TAKTDVNGLA-TFDLKSSKQEDNTVETILENGVKQTLIVSPGDSSTAQVDLOKSK 848  
 OY 113 RELQIKGY-KOLLEKSVNCPGSLTTLNSAHFNCKRNAASGASLYIYPAGELNLPFGGIW 171  
 Db 849 NEVVADGNDGATMTATVBDKAGNLDVKTFFVNSAAKLSQTEVNSHD-----GIA 901  
 OY 172 DATLKLRVRRRYSETGTYTINTIKLTDKGNIQIWLPOKRSARVDLNRPTGGGYIG 231  
 Db 902 TATLT-----SLKNGDYVTVASVSSGQANQOVIFIGDQSTALTLVS-PSGDITV-- 951  
 OY 232 RNSVDMCFYDGYSTNSSLE---IRFODNPNKSDCKEFLRKINDDTKEIAYTLISLLAG 287  
 Db 952 -----TNTAPLHMTATLQDKNPNPKDKKEITFSVND-----VASFFSISNSG 994  
 OY 288 KSLTPINGTSL-----NIADAASLETNNRITAVTMEISVPLC 327  
 Db 995 KGMTDSNGTAIASLTGLAGTHMITARLANSVSDQPMTFVADKRAVVAVLQTSKAEII 1054  
 OY 328 WPG--RLQDRAKVENPEAGQYMGINIVTTPS 357  
 Db 1055 GNGVDETTTLTATVKDP-FDNVKNLSVVERFS 1085

## RESULT 7

S22835

alpha-agglutinin - yeast (Saccharomyces cerevisiae)  
 N/Alternate names: 22k glycoprotein; protein J1418; protein YJR004c

C/Species: Saccharomyces cerevisiae

C/Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 29-Oct-1999

C/Accession: S22835; S51229; A32822; S55192; S57019

R:Husser, K.; Tanner, W.

FEBS Lett. 255, 290-294, 1989

A/Title: Purification of the inducible alpha-agglutinin of S. cerevisiae and molecula

A/Reference number: S22835; MUID:90005993; PMID:2676603

A/Accession: S22835

A:Molecule type: DNA

A:Residues: 1-650 <HAY>

A/Cross-references: EMBL:X16861; NID:g33552; PIDN:CAA3752.1; PID:g3353

A/Accession: S51229

A:Molecule type: protein

A:Residues: 20-24 <HA2>

R:Lipke, P.N.; Wojciechowski, D.; Kurjan, J.

Mol. Cell. Biol. 9, 3155-3165, 1989

A/Title: AG-alpha-1 is the structural gene for the Saccharomyces cerevisiae alpha-299

A/Reference number: A32822; MUID:90014768; PMID:2677666

A/Accession: A32822

A:Molecule type: DNA

A:Residues: 1-448, 'P', 450-555, 'E', 557-580, 'L', 582-650 <LIP>

A/Cross-references: GB:M28164; NID:g171041; PIDN:AAA34417.1; PID:g171044

R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A/Reference number: S55183

A:Molecule type: DNA  
 A:Residues: 1-650 <DEH>  
 A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577  
 Ride Haan, M.; Griwell, L.A.; Smits, P.H.M.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56771  
 A:Accession: S57019  
 A:Molecule type: DNA  
 A:Residues: 1-650 <ZAG>  
 A:Cross-references: EMBL:249504; NID:g1015625; PIDN:CAA69526.1; PID:g1015626; MIPS:YUR00  
 C:Genetics:  
 A:Gene: SGD:SAG1; AGAL1; AGALPHA1  
 A:Cross-references: SGD:S0003764; MIPS:YJR004c  
 A:Map position: 10R  
 C:Keywords: glycoprotein

Query Match  
 Best Local Similarity 20.9%; Score 114; DB 2; Length 650;  
 Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

```

4 ILFIPLFFSVL-----FTFA-----VSADKIPG-----DESITNIFGPRDRNE--- 43
8 ILMLRSLALASAININDITFSNLEITPLANKQPOGWTATPFDLSIADASSIREDEFTL 67
44 SSP---KANILNNHTA-----YSESHLYDRMTLC-----LSHNTLN 80
68 SMPHYRIKLINNSQTATISLADGTEAFKCYVSOQAAYLYENTFTCTAQNDSYSTNTD 127
81 GACPTSENPSSSVGETNITLQTEKRSILKRELIQINGKQLLRKSYNC-PSGLTLNSA 139
128 GSITSLNFDGGSSEYEL-----ENAKFKSGMLVKLGOMSDVNVNEDTAETENF 182
140 HFNCKNNAASGASLYL-YIPAGELKNLPFGGIMDATILKRYKRRYSEYGTGTINTITIK 197
183 HSGRSTGYGSPFSEYHLYGKCPNGYF---LGS---TEKI-----DYSSNNNVLD- 225
198 LTKDNINQI-----WLPQKSDARVDLNLPTGGGTYGRN---SVMCKFYDG--- 242
226 --DCSSVQVYSSNDENDMWFPOSYNDTNADV-----TCGSMWMTLDEKLYDGEML 275
243 -----YSTNSSSLE--TRFODNN--PKSDCKF 265
276 WYNALQSLPANVNTIDHLEFOYTOLODTIANTTYATOFSTIREFIVYGRNLGTSASAKS 335
286 YLRKINDTKETI--AYLS-----LLAGKSLRPTNGSLNIADAA 304
336 FISTTTDLTSLINTSAYSTGSIYETGNRTTSEVISHVVTSTKLSPATSTLTIAQTS 395
305 SLETFWN-----RITAVMPETISVPYLCPPGR 332
396 IYSTDNITVGTDIHTTSEVSDVETISRETASTVVAAPTSTGTGTAM 444

```

Db

RESULT 8  
 D64962  
 probable membrane protein b1978 - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C:Accession: D64962  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: D64962  
 A:Molecule type: DNA  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Residues: 1-2383 <BLAT>  
 A:Cross-references: GB:A000289; GB:U00096; NID:g1788285; PIDN:AAC75042.1; PID:g1788288;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Keywords: nucleotide binding; P-loop; transmembrane protein  
 F:54-70/Domain: transmembrane #status predicted <TM>  
 F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match  
 Best Local Similarity 5.8%; Score 109; DB 2; Length 2383;  
 Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

```

13 SSVLETFPAVSADKIPGDSITNIFGPRDRNESSPKHNLNNHTAY---SESHLYDRMT 69
762 SAKIATLSASNNGVANNAATVSVNVADEGS---NPINHVTFAVLSSASNSFNQ 818
70 FLCLSHNTLNCACPTSENPSSSVGETNITLQTEKRSIL-----K 112
819 ----TAKIDVNGLA-TFDLKSSKQEDNVEYTLLENGVQTLIVSGSSRAVDLQSK 873
113 RELQIKYKQUL-LFKSVNCPGSLTNSAHFNCKNNAASGASLYLYIPAGELKNLPFGGIW 171
874 NEVVADGNDVMTATFATVDAGKNLNDVMTFENVASAKLSQTEVNSHD-----GIA 926
172 DATLKLRYKRRYSEYGTITNTITKLDKNGIQIWLPOFKSDARVDLNLPTGGGTYIG 231
927 TATLT-----SLKNGDYRYTASVSSGSOANOQVNFIGDSTALTLTSV-PSGDITV-- 976
232 RNSVDMCFYDGYSTNSS---SLEIRFODNPKSGKRYLKRINDDTKEIAYTSLIAG 287
977 -----TTFAPQYMATILODKNGNPKDKETTFVSPND---VASKTSISNGG 1019
288 KSLPTNGTSL-----NIADAASLETNWRITAVTPEISVPYLC 327
1020 KGMTDSNGVALIASLGLTAGTHIMARLANSVSDAQMTFVADKRAVVVYLQTSKAEII 1079
328 WPG--RLQIDAKVENPEAGQTMGINVTFT 355
1080 GNGVDETTLTATVKDP-SNHVPAGITVNT 1108

```

Db

RESULT 9  
 S55097  
 probable membrane protein YMR215w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein YMR261.09  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
 A:Reference number: S55097  
 A:Accession: S55097  
 A:Molecule type: DNA  
 A:Residues: 1-524 <DED>  
 A:Cross-references: EMBL:249809; NID:g854459; PID:g854467; GSPDB:GN00013; MIPS:YMR215  
 A:Experimental source: strain AB972  
 C:Genetics:  
 A:Gene: MIPS:YMR215w  
 A:Cross-references: SGD:S0004828  
 A:Map position: 13R  
 C:Superfamily: glycopospholipid-anchored surface glycoprotein GASI  
 C:Keywords: transmembrane protein  
 F:7-23/Domain: transmembrane #status predicted <TM>

Query Match  
 Best Local Similarity 5.7%; Score 108; DB 2; Length 524;  
 Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

```

119 GYKUL-LFKSVNCPGSLTNSAHFNCKNNA--SGASLYLYIPAGELKNLPFGGIWDA 173
264 GYDKLNSTFEDAVIP---LIFEYGCNKNRPREFDEVSEGLY---GGLKNVFGSLT- 313
174 TLKLRYKRRYSEYGTITNTITKLDKNGIQIWLPOFKSDARVDLNLPTGGGCT 228
314 -----VEYTEEANNVGL---VKLDSGSL-----TKDQFVWLESQLKVNSLPTKES 359
229 YIGRNSVDMC-----FYDGYSTNSSSLE-----IRFODNPKSDSK----- 264
360 EISDSIYKCNMSATINISGFGTNNFTLPSQPAEINMIIEYGVNGTNT-GIILTVAVP 418

```

Db



QY 265 ----FYLRKINDTKELAVYLLSLAGKSLPTNGTSLNADASLETNMR-ITAVTMP 319  
 DB 419 TFFNTYIKNNKQDF--ISATISYDKA-NSLNELDYATATYAKASNTSSQSSRSLTSTSP 475  
 QY 320 EISVPLVPCWPGRLQDLAKVENPAGQYMGNT-NVTFPPS 357  
 DB 476 SSGTSSSSSTG-----SSASSSSKSGKGVNIVVSPSQS 510

## RESULT 10

B64635

toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C&gt;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999

C:Accession: B64635

R:omb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khatik, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. Reference number: A64520; MID:97394467; PMID:9252185

C:Accession: B64635

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2529 &lt;TOM&gt;

A:Cross-references: GB:AE000602; GB:AE000511; MID:92314060; PIDN:AAD07969.1; PID:9231406

Query Match 5.7%; Score 108; DB 2; Length 2529;  
 Best Local Similarity 22.5%; Pred. No. 16;  
 Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENPSSSVSGEINTILOFTEKRSLL-----KRELOIKGYKOLLF--SVNCP 131  
 DB 977 TSNFNTTQGLGNTNTL---SSQSLNFGNTOITLNNANITLGNKSQAFFKSLTLDNN 1033  
 QY 132 SGLTL-NSAFNCKNNA--SGASLYLYPA-GELKNLPFGIMDATLKLVRKRYSEY 187  
 DB 1034 SNLSLDSQVNLNANNNAFNNQASLNLYNGSQTFNSLFENG--GTLISLASSKLNAS 1090  
 QY 188 GRYTITITIKLPD-----KCNIQIWLPGFSDARVDNLNLPFGGCTYIGNSVD 236  
 DB 1091 AFSNNNTTINLDSVLSASNTSLNNT-----NFGASQADP-----GNTITIDYASFN 1140  
 QY 237 MCFYDYSYSSSL-----EIRFQDNPK-----SDGKFLRKIN--DDP 274  
 DB 1141 --FDSASSLNPNLNTANGALNFNGYTPSLTKALMSVSGQVGLGNNNDINLSDINITDNT 1198  
 QY 275 KEIAYVLSLLAGKSLPTNGT-----LNIDAAASLETN---WNRTAVTMPET 321  
 DB 1199 KSVTY--NIIINAKGIGTIGSANGYKILFYCGMKIQAATYSDNNNIQTWSFNPINSQI 1256  
 QY 322 -----SVVYL-----CMPGRQLQDAKYENP-----EAGQY-M 347  
 DB 1257 IQESINGDLTTEVLNPNNSASNTIFNIAPELYNQAQKONPGYSYDYSDNAGTYIYL 1316  
 QY 348 GNINVTFT-SSQT 360  
 DB 1317 SNIKGLFTPKGSQT 1330

## RESULT 11

A64904

probable fimbrial protein b1502 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C&gt;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: A64904

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MID:97426617; PMID:9278503

A:Accession: A64904

A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-304 <BLAT>  
 A:Cross-references: GB:AE000247; GB:U00096; MID:91787773; PIDN:AAC74575.1; PID:917877  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Superfamily: fimbrial protein f1mH  
 C:Keywords: fimbria

Query Match 5.7%; Score 107; DB 2; Length 304;  
 Best Local Similarity 21.0%; Pred. No. 1.2;  
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IFTLFFSVLFTFAVSADKIGDESI-----TNIFGRPRNESPKNHNT---LNH 54  
 DB 8 KVLFGIYLLMAKVAFAFSCVND---GGSSIGAGTISVYNLD-PIYIQGNLYVDLSQH 63  
 QY 55 ITAVSSHLYDRMTFLCLSSH-NTINGACPTSENPSSSVSGEINTILOFTEKRSLLIKR 113  
 DB 64 ISCMNDYGGWYD-----TDHINLVQ-----SAFAG----- 89  
 QY 114 ELQIKGYK-QLFKSYNCPGSLNLSAHFNCKNNAASGASLYLYIPAGELKNLPFGIMD 172  
 DB 90 --SLQSYKSGSLYNNNTYPPPLTTNTNVLDIGDKTPMPLPKLYI-----TPVGAAG 140  
 QY 173 ATLK-----LVRKRRRYSEYGT-----YTINITIK-----LTDKGNIOIWLPGFK 212  
 DB 141 VVIRKAGEVARIMHYKATILGSGNPNRFTNIIISNNNVMPPTGCGCTVDSRNTVVDLPDF 200  
 QY 213 SDARVDNLNLPFGGCTYIGNSVDKCFYDQSTSNSSLEIRPDNNPKSGKRYLKRIND 272  
 DB 201 GSNEIPL-----GYVCS-SEQKLSFYLSGATDSSROY-FANTAP-----D 239  
 QY 273 DFEIAYTSLLAGK-----SLPTNGTSLNADASLETNMRITAVTMPET 321  
 DB 240 ATNASGVGTILMNGKILATGENVSLGTAKSKVPLGLSATYQGTGNKVSAGTYQSV 296

## RESULT 12

C90892

probable adhesin [imported] - Escherichia coli (strain 0157:H7, substrain RMD 050995)

C:Species: Escherichia coli

C&gt;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: C90892

R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g A:Reference number: A93629; MID:21156231; PMID:11238796

A:Accession: C90892

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 &lt;HAY&gt;

A:Cross-references: GB:BA000007; PIDN:BA035530.1; PID:913361573; GSPDB:GN00154

A:Experimental source: strain 0157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECS2107

C:Superfamily: fimbrial protein f1mH

Query Match 5.7%; Score 107; DB 2; Length 304;  
 Best Local Similarity 21.0%; Pred. No. 1.2;  
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IFTLFFSVLFTFAVSADKIGDESI-----TNIFGRPRNESPKNHNT---LNH 54  
 DB 8 KVLFGIYLLMAKVAFAFSCVND---GGSSIGAGTISVYNLD-PIYIQGNLYVDLSQH 63  
 QY 55 ITAVSSHLYDRMTFLCLSSH-NTINGACPTSENPSSSVSGEINTILOFTEKRSLLIKR 113  
 DB 64 ISCMNDYGGWYD-----TDHINLVQ-----SAFAG----- 89  
 QY 114 ELQIKGYK-QLFKSYNCPGSLNLSAHFNCKNNAASGASLYLYIPAGELKNLPFGIMD 172  
 DB 90 --SLQSYKSGSLYNNNTYPPPLTTNTNVLDIGDKTPMPLPKLYI-----TPVGAAG 140



```
OY 210 OFKSDARVDJN-----LRPT-----GGTYIGRNSVDMCFYDGYSTNSSLE----- 251
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 -FIKGNIDISGOSALLRPSNITSTFGHNGNIDINTLNLISGGGSSSTLSAKA 530
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 252 -----IRFODNNPK--SDGFYLRKINDPTKEIAYTLLSLLAGKS----- 289
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 531 GNISINSSNINVTGNINSNSPFSFINSNELL--VDPNLQKLLYRQPPLLIGQACNIFL 588
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 290 -----LFTNGTSLN-----IADAASLETMMNRITAVTPEISVPVLCWP-GRQLDAKVE 339
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 589 NTDIINISNGGLINARNEGVDAGNIRISANTINISOGEVNATYTTIGEGNIIILNSR-- 646
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 340 NPEAGQYMGNIIVTFT 355
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 647 -----NLFMNSRTTAT 658
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Search completed: August 19, 2003, 09:49:00  
Job time : 46 secs

**This Page Blank (uspto)**



QY 121 KOLLEKSVNCPBGLTNSAHFNCNNAASGASLYIYPAGELNLPFGIMDATLRLVK 180  
 DB 121 KOLLEKSVNCPBGLTNSAHFNCNNAASGASLYIYPAGELNLPFGIMDATLRLVK 180  
 QY 181 RRYSEYGVYNTINTIKTLKDKNGIOIWLPOFKSDARVNLNLPRTGGGYIGNSVDKCFY 240  
 DB 181 RRYSEYGVYNTINTIKTLKDKNGIOIWLPOFKSDARVNLNLPRTGGGYIGNSVDKCFY 240  
 QY 241 DGYSTNSSLEIRFODNNPKSPDKFYLRKINDTKEIAYVTLISLAGSLPPTNGSTINI 300  
 DB 241 DGYSTNSSLEIRFODNNPKSPDKFYLRKINDTKEIAYVTLISLAGSLPPTNGSTINI 300  
 QY 301 ADASLEFNWNRITAVTPEISVPLCPWGLQDLDAKVENPAGYMGYNINVTFTPSST 360  
 DB 301 ADASLEFNWNRITAVTPEISVPLCPWGLQDLDAKVENPAGYMGYNINVTFTPSST 360  
 QY 361 L 361  
 DB 361 L 361

## RESULT 2

YEEI\_ECO57 STANDARD: PRT: 2660 AA.

AC Q8XSV7: Q8X2B9: Q8X2C0: 2660 AA.  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yeeI.  
 GN 23135 OR EC52775/EC52776.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE-21074935: PubMed-11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postal G., Hayek J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.,  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RT Nature 409:529-533(2001).  
 RL [2]

## SEQUENCE FROM N.A.

STRAIN-O157:H7 / RMD 050952;  
 MEDLINE-21156231: PubMed-11238796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohnishi M., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kubera S., Shiba T., Hattori M., Shitagawa H.,  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -1 SIMILARITY: Contains 16 B1q-1 domains.  
 CC -1 SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.  
 CC -1 CAUTION: Ref.2 sequence differs from that shown due to a  
 CC frameshift in position 1315.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE005423: AAC57041.1: -  
 DR EMBL: AP002559: BAB36198.1: ALT\_FRAME.  
 DR EMBL: AP002559: BAB36199.1: ALT\_FRAME.

DR InterPro: IPR003344; BIG\_1.  
 DR InterPro: IPR003353; Intimin.  
 DR InterPro: IPR006601; PKD\_domain.  
 DR Pfam: PF02369; BIG\_1; 16.  
 DR PRINTS: PR01369; INTIMIN.  
 DR SMART: SM00634; BIG\_1; 16.  
 DR SMART: SM00089; PKD; 8.  
 KW Hypothetical protein; Repeat; Complete proteome.  
 FT DOMAIN 738 834 BIG-1.1.  
 FT DOMAIN 840 929 BIG-1.2.  
 FT DOMAIN 931 1033 BIG-1.3.  
 FT DOMAIN 1042 1132 BIG-1.4.  
 FT DOMAIN 1134 1236 BIG-1.5.  
 FT DOMAIN 1245 1335 BIG-1.6.  
 FT DOMAIN 1337 1439 BIG-1.7.  
 FT DOMAIN 1448 1539 BIG-1.8.  
 FT DOMAIN 1548 1652 BIG-1.9.  
 FT DOMAIN 1653 1750 BIG-1.10.  
 FT DOMAIN 1751 1855 BIG-1.11.  
 FT DOMAIN 1856 1957 BIG-1.12.  
 FT DOMAIN 1963 2056 BIG-1.13.  
 FT DOMAIN 2065 2156 BIG-1.14.  
 FT DOMAIN 2157 2252 BIG-1.15.  
 FT DOMAIN 2254 2355 BIG-1.16.  
 SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 6.4%; Score 120; DB 1; Length 2660;

Best Local Similarity 20.4%; Pred. No. 0.62;  
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

QY 13 SSVLFTFAVSADKIPGDSITINIGPRDRNESSPKHNLNHNITAY---SESHLYDPM 69  
 DB 737 SAKIATLSASNNGVLANENANVTAVSVNADREGS---NPINHTVFAVLSSGASFSNNQ 793  
 QY 70 FLCSSHMLNGACPTSSNPSSSVSGFNITLOFTEKRSIL-----K 112  
 DB 794 ----TAKTDVNGLA-TPLDKSSKQEDNTVEVTLNGVQOTLVSVGDSSTQAQVDLQSK 848  
 QY 113 RELQIKGY-KOLLEKSVNCPBGLTNSAHFNCNNAASGASLYIYPAGELNLPFGIM 171  
 DB 849 NEVAVADGSDASTMTFTVDADGNLNDVKYFENVVSAAKLSQFEVNSHD-----GIA 901  
 QY 172 DATLKLRYKRRYSEYGYTNTIKTLKDKNGIOIWLPOFKSDARVNLNLPRTGGGYIG 231  
 DB 902 TATLR-----SLKNGDYTVASVSSGQANQVITFGDSTALTLSTV-PSGDIIV-- 951  
 QY 232 RNSVDMCFYDGYSTNSSLE-----IRFODNNPKSPDKFYLRKINDTKEIAYVTLISLAG 287  
 DB 952 ----TNTAPLHMTATLQDKNGNPKDKKEITFSVPND---VASRSISNSG 994  
 QY 288 KSLPPTNGTSL-----NIADAASLEFNWNRITAVTPEISVPLIC 327  
 DB 995 KGMIDSNCTALASLGLTAGHMITARLANSVSDTQMTFPAADRANVYLQTSKELI 1054  
 QY 328 WPG--RLQDLAKVENPAGGYMGYNINVTFTPSST 357  
 DB 1055 GNGVDETLTFTATVKDP-EDNVVYKNLISVVERIS 1085

## RESULT 3

SAGI\_YEAST STANDARD: PRT: 650 AA.

AC P20840:  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Alpha-1-galactosidase precursor (Ag-alpha-1).  
 GN SAGI OR AGAL1 OR YJR004C OR J1418.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]





ID	YEEJ_ECOLI	STANDARD:	PRT: 2358 AA.
AC	P76347; P94750;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Hypothetical protein yeeJ.		
GN	yeeJ OR B1978.		
OS	Escherichia coli.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RC	MEDLINE=97426617; PubMed=9278503;		
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Pena N.T., Burland V.,		
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,		
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,		
RA	Mau B., Shao Y.;		
RL	"The complete genome sequence of <i>Escherichia coli</i> K-12.";		
RL	Science 277:1453-1474(1997).		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12;		
RC	MEDLINE=97251358; PubMed=9097040;		
RA	Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,		
RA	Iseno K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,		
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,		
RA	Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,		
RA	Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,		
RA	Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;		
RT	"A 460-kb DNA sequence of the <i>Escherichia coli</i> K-12 genome		
RT	corresponding to the 40.1-50.0 min region on the linkage map.";		
RL	DNA Res. 3:379-392(1996).		
CC	-1- SIMILARITY: Contains 13 Btg-1 domains.		
CC	-1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; AE000289; AAC75042.1; ALT_INIT.		
CC	EMBL; D90837; BAAL5800.1; -		
CC	EMBL; D90836; BAAL5799.1; ALT_INIT.		
CC	Ecogene; EGI3378; yeeJ.		
DR	InterPro; IPR003344; Btg_1.		
DR	InterPro; IPR003535; Intimin.		
DR	InterPro; IPR002482; LysM.		
DR	InterPro; IPR000601; PKD_domain.		
DR	Pfam; PF02369; Btg_1; 13.		
DR	PRINTS; PRO1369; INTIMIN.		
DR	SMART; SMO0634; BID_1; 13.		
DR	SMART; SMO0257; LysM; 1.		
DR	SMART; SMO0089; PKD; 6.		
KW	Hypothetical protein; Repeat; Complete proteome.		
FT	DOMAIN 788 834 Btg-1.1.		
FT	DOMAIN 840 931 Btg-1.2.		
FT	DOMAIN 932 1033 Btg-1.3.		
FT	DOMAIN 1042 1137 Btg-1.4.		
FT	DOMAIN 1146 1237 Btg-1.5.		
FT	DOMAIN 1246 1350 Btg-1.6.		
FT	DOMAIN 1351 1448 Btg-1.7.		
FT	DOMAIN 1449 1553 Btg-1.8.		
FT	DOMAIN 1554 1655 Btg-1.9.		
FT	DOMAIN 1661 1754 Btg-1.10.		
FT	DOMAIN 1763 1853 Btg-1.11.		
FT	DOMAIN 1855 1950 Btg-1.12.		
FT	DOMAIN 1952 2053 Btg-1.13.		

FT	CONFLICT	105	105	S -> G (IN REF. 2).
SO	SEQUENCE	2358	AA: 248599	MM: 23224975056F631ED CRC64;
	Query Match		5.8%, Score 109;	DB 1; Length 2358;
	Best Local Similarity	20.0%,	Pred. No. 3.7;	
	Matches	78;	Conservative	65; Mismatches 157; Indels 90; Gaps 15;
QY		13	SVSLFETFAVSADKLPGEDESTINICOPDRNRSSPKHNLTNNHTAY---	SSSHLYDGMT 69
DB		737	SAKATATLSASNNGLANENANNTYVSNAVADGS---	NFINHTYTFVTLSSSAISFNQN 793
QY		70	FLCLSSHNTLNGACPTSENPSSSSVSGEFTNTLTQTEKRSLLI-----	K 112
DB		794	---TAKTDVNGLA-TFDLKSSKQEDNTVEVTELTENGWYQTLTIVSFVGDSSPTAOYDLQSK 848	
QY		113	RELQIKGYKQL-LKSYNCPBGLTLNSHFHCNNMAASGASLYITPAGELKNLPFGIWM 171	
DB		849	NEVAVADNDSTYMTATYATADAKGNLLNLYMTFVFNNSAEAKLSQTEVNSHD-----	GIA 901
QY		172	DATLKLKRRYSSEYGYTNTITIKTLKDNQIOIMLPQFESDARVDLTLPRTGGTYIG 231	
DB		902	TATLT-----SLKNDGYRTATSVSSSSQANQVNFIDGSTALTALTSV-PSGDITY--	951
QY		232	RNSVDMCFYDGYSTNSS---SLEIRQDNNPKSDGKFYLRKINDDTKEALNTYLSLLIAG 287	
DB		952	-----TNTAPQYWTATLQDKNPNLKDKEITFESVPD---VASKSISISNG 994	
QY		268	KSLEPTNGTSL-----NINDAASLENNRITAVTWPETISVPYC 327	
DB		995	KGMTDSGVATLASLTGLTLAGTHMARIANSVNSDAQPMFTVADKRAVVLQTSKAEIT 1054	
QY		328	WPG--RLQLDAKVENPEAGQYMGMINVYFT 355	
DB		1055	GNGVDEFTLLTFATVKDP-SNHPVAGITVNET 1083	
	RESULT 5			
	GAS3_YEAST			
ID	GAS3_YEAST	STANDARD:	PRT:	524 AA.
AC	003655;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	GAS3 protein precursor.			
GN	GAS3 OR YMR215W OR YMR261.09.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=5288C / AB972;			
RX	PubMed=9169872;			
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,			
RA	Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,			
RA	Jagels K., Iyer G., Moule S., Odell C., Pearson D., Rajandream M.A.,			
RA	Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.G.,			
RT	"The nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	XIII."			
RL	Nature 387:90-93(1997).			
RN	{2}			
RP	IDENTIFICATION.			
RP	MEDLINE=20529944; PubMed=11079560;			
RA	Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gil C.,			
RA	Nombela C.;			
RT	"A proteomic approach for the study of Saccharomyces cerevisiae cell			
RT	wall biogenesis."			
RL	Electrophoresis 21:3396-3410(2000).			
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(Potential)			
CC	-I- SIMILARITY: Belongs to the GAS1 family.			
CC	-----			
CC	This SWISS-PROT entry is copyright It is produced through a collaboration			

RP SEQUENCE FROM N.A.;  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Db 240 ATKASGVGVTLMRNGKILATGENVSLGTVNKS KVP LGLSATYGGTGNKVSAGTVQSV 29

DI 13-SEP-2003 (REL. 4Z, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Autolysin precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)  
 DE (Beta-glycosidase).  
 GN EF0799.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91358349; PubMed=1679432;  
 RA Bellevue C., Potvin C., Trudel J., Asselin A., Bellemare G.;  
 RT "Cloning, sequencing, and expression in *Escherichia coli* of a  
 RT Streptococcus faecalis autolysin";  
 RL J. Bacteriol. 173:5619-5623(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,  
 RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,  
 RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouli H.,  
 RA Ueberbach T., Radue D., Ketchum K.A., Dougherty B.A., Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis";  
 RL Science 299:2071-2074(2003).  
 CC -1- FUNCTION: Hydrolyzes the cell wall of *E. faecalis* and  
 CC M. lysodeikticus. May play an important role in cell wall growth  
 CC and cell separation.  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN  
 CC BINDING.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.  
 CC -1- SIMILARITY: Contains 6 LysM repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, M58002; AAA67325.1; -;  
 DR EMBL, AE016949; AA080613.1; -;  
 DR PIR, A38109; A38109.  
 TRIGR, EF0799; -;  
 DR InterPro: IPR002901; Amidase\_4.  
 DR InterPro: IPR002482; LysM.  
 DR Pfam: PF01832; Amidase\_4; 1.  
 DR Pfam: PF01476; LysM; 5.  
 DR SMART: SM00257; LysM; 6.  
 DR SMART: SM00047; LY22; 1.  
 DR Hydrolyase: Glycosidase; Bacteriolytic enzyme; Cell wall;  
 KM Cell division; Septation; Repeat; Signal.  
 FT SIGNAL 1 53  
 FT CHAIN 54 737  
 FT REPEAT 363 405  
 FT REPEAT 431 473  
 FT REPEAT 499 541  
 FT REPEAT 567 609  
 FT REPEAT 633 675  
 FT REPEAT 695 737  
 FT REPEAT 85 737  
 FT CONFLICT 118 85  
 FT CONFLICT 143 118  
 FT CONFLICT 417 417  
 FT CONFLICT 449 449  
 FT CONFLICT 476 476  
 FT CONFLICT 484 484  
 FT CONFLICT 567 632  
 FT SEQUENCE 737 AA; 77025 MW; ABB16BD506AC7507 CRC64;

Query Match 5.3%; Score 100; DB 1; Length 737;  
 Best local Similarity 22.9%; Pred. No. 4.1;  
 Matches 69; Conservative 39; Mismatches 115; Indels 78; Gaps 15;  
 QY 41 RNESSPRKNI-LNNHITAVSESHLYDRMTFLCLSHNTLNGACPTSPSSSSV----- 94  
 DB 309 RYATDPSPYAKLNINVTAY--NLTYDFPSSGNTGGTGVNPGTGSNNQSGTNYTYVK 366  
 QY 95 SGET--NITLQF-----TEKRSI--IKRELQINGYQQLFKSWNCPSGLTLNNAHRCNKN 146  
 DB 367 SDDTLNKKIAAQYGVSVANLRSMNGISGDLIFVQKLIYKKA--SGNTGGSGSGSN 423  
 QY 147 AAGSASLYIPAGELKNLPFGIWDATLKLVRKRSYSETGYTITINIKTLDKGNIOI 206  
 DB 424 -QSGINTYTYVKSGLTN-----KIAAQY-----VSANLR 455  
 QY 207 WLQPKSDARVDLNRPTGGGTYIGRNSVDMCFYGYSTNSSLEIRPDNNPKSDGFY 266  
 DB 456 W-----NGISGDL-----IFVQK--LIVKKGASGNTGSGNNGSGNTGYTY 499  
 QY 267 LKINDPKEIA--YTLG-----LLAGKSLTPNGSLNINADAALETNN 311  
 DB 500 TTKSGDTLKKIAAQYGVSVANLRSMNGISGDLIFVQKLIYKKGTSNTGSGSGSN 559  
 QY 312 R 312  
 DB 560 Q 560  
 RESULT 8  
 SYFR, LISIN  
 ID SYFR, LISIN STANDARD; PRT; 802 AA.  
 AC 092C16;  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)  
 DE (Phenylalanyl-tRNA ligase beta chain) (Phers).  
 GN PHER OR L1N1185.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangoul L., Buchrieser C., Rusnijk C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Chardit A., Chelouani F., Couve E., de Baurvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussuguet O.,  
 RA Entlian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,  
 RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjati H.,  
 RA Nordstiek G., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schueter T., Simoes N., Tietz A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species";  
 RL Science 294:849-852(2001).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
 CC diphosphate + L-phenylalanyl-tRNA(Phe).  
 CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).  
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain  
 CC family. Subfamily 1.  
 CC -1- SIMILARITY: Contains 1 tRNA-binding domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AL596167; CAC96416.1; -  
DR PIR: AH1580; AH1580.  
DR Listlist: LIND1648; -  
DR HAMAP: MF\_00283; -; 1.  
DR InterPro: IPR005146; B3\_4.  
DR InterPro: IPR005147; B5.  
DR InterPro: IPR005121; Fdx-AnticB.  
DR InterPro: IPR004532; Phet\_bact.  
DR InterPro: IPR002547; trna\_bind.  
DR Pfam: PF03483; B3\_4; 1.  
DR Pfam: PF03484; B5; 1.  
DR Pfam: PF03147; FDX-ACB; 1.  
DR Pfam: PF01388; trna\_bind; 1.  
DR TIGRFAMs: TIGR00472; Phet\_bact; 1.  
DR PROSITE: PS50886; TRBD; 1.  
KM Aminocyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
Metal-binding; Magnesium; RNA-binding; trna-binding;  
Complete proteome.  
FT DOMAIN 40 155 TRNA-BINDING.  
FT METAL 462 462 MAGNESIUM (BY SIMILARITY).  
FT METAL 468 468 MAGNESIUM (VIA CARBONYL OXYGEN) (BY  
SIMILARITY).  
FT METAL 471 471 MAGNESIUM (BY SIMILARITY).  
FT METAL 472 472 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 802 AA; 86191 MW; D7560219364D7D0 CRC64;

Query Match 5.2%; Score 99; DB 1; Length 802;  
Best Local Similarity 17.3%; Pred. No. 5.4;  
Matches 70; Conservative 63; Mismatches 132; Indels 140; Gaps 15;

QY 48 HNLNHNITAYSSSHLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNTTQFTK 107  
DB 422 NRIIGTEI-SLSEIETIFRLGFLVEKEDTLIEVPTF-----RMDITIE----- 466  
QY 108 RSLIKRELQKKGKOLLFFSVNPSGLTNSAHFNCKNAASGLIYIPAGELKNLF 167  
DB 467 ADILEVARIYIGDEL---PVLPATST----- 492  
QY 168 GGIWDATLTKRVKRYSETYG-----TYIN-----ITIKLFDKNGIQIWLPOFKSDARY 217  
DB 493 GGSIDSQKARVRWALLEGAGLNLQALTYLTSKKDTRLALDEKVALSMPSSESHL 552  
QY 218 DLNLRP---TGGGTYIGRNSVDMCFYD---GYSTNSSLEIRFO----- 255  
DB 553 RTSIVPOLIRASYNIRAKRMVALYEMGTVFYATEGDMLPTEQHLAQLTGCMHTADM 612  
256 DNNPKS-----DGKYLKINDTELEIAYTSLSLAGK-----SL 290  
DB 613 QKTPKRVDFVKGIVGLVNRKLGIEALHWKQIEEELHPRTASTIOGKEIGYGLAL 672  
QY 291 TPTNGSLNIADAASLETWMNRITAVTMEISY-PVLCWP----- 329  
DB 673 HRAVEASYDLKETVEINVKALDTEKRYVYHPKPKPEMTROLALLVDKDTHTATIS 732  
QY 330 -----GRLOLDAKVENPEAOQYMGN-----INVFTPPSSOTL 361  
DB 733 OVYKEHGNLVDIELDFEGESLGENKSLAYLTFLDSERTL 777  
RESULT 9  
IGA4\_HAEIN  
ID IGA4\_HAEIN STANDARD; PRT; 1849 AA.  
AC P45386;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
GN IGA.

OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBL\_taxid=27;

RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-NHTI HK61;  
RX MEDLINE-92234949; PubMed-1373717;  
RA Poulsen K., Reinholdt J., Kilian M.,  
RT "A comparative genetic study of serologically distinct Haemophilus  
influenzae type I immunoglobulin A1 proteases."  
RL J. Bacteriol. 174:2913-2921(1992).  
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
PRODUCING INTACT FC AND FAB FRAGMENTS.  
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
certain Pro-Xaa bonds in the hinge region. No small molecule  
substrates are known.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M87491; AAA24968.1; -  
DR MEROPS: S06.001; -  
DR InterPro: IPR006315; Autolysinport.  
DR InterPro: IPR005546; Autolysinport.  
DR InterPro: IPR000710; IGA\_S6.  
DR InterPro: IPR004899; Pertactin.  
DR Pfam: PF03797; Autolysinporter; 1.  
DR Pfam: PF02395; IGA1; 1.  
DR Pfam: PF02312; Pertactin; 1.  
DR PRINTS: PR00921; IGASERPRASE.  
DR TIGRFAMs: TIGR01414; autolysin\_bar1; 1.  
KW Hydrolyase; Serine protease; Transmembrane; zymogen; signal.  
FT SIGNAL 1 25  
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.  
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).  
FT ACT\_SITE 299 299 PROBABLE.  
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 5.2%; Score 99; DB 1; Length 1849;  
Best Local Similarity 23.4%; Pred. No. 16;  
Matches 76; Conservative 56; Mismatches 127; Indels 66; Gaps 20;

QY 19 FAVSADKIGDSDSITNIFGPRDRNESSPHNLIHNITAYSSS---HTLVDKMTFLCLSS- 75  
DB 754 FKATIMNVGNASLVS-----GRNVA-----NITSN-TTASNAQAIIIGKTDGFCVSRSD 803  
QY 76 -----NHT-LNGACPTSENPSSSVSGETNIT---LQFTKRSILIKRELQIKGYKQLLF 125  
DB 804 YNGVYTCNHSNLSERKALNSFN---TNLRKNVNLTEASTLTKANLFGTIOGIGSQVNL 861  
QY 126 KSYNCPDGLTNSAHFNCK-NAASGASLYIYPAGELKNLDPGGIWDATLTKRVKRRYS 184  
DB 862 KE-NSHWMLTGNS---NVNQLNLFTNG---HILNAQNNDAN-----KV 896  
QY 185 ETVGYTYTINITIKLIDKNGIQIWLPOFKSDA-RVLDNLNLRPGGGTY-----IGR-NSVDM 237  
DB 897 TTYNTLTVN---SLSGNSGFYIWDFTNKKSKVYVNSATGNTFLQYADKTGEPRHNL 953  
QY 238 CFYDGYSTNSSLEIRFODNN-PSKDGKREYLRKINDTELEIAYTSLSLAGKSLTPNTGT 296

```

Db      954 TLEFASNATRNNEVTLANGSVDRGAMKKLRNVNG--RYDLNVEVERKNOTVDITNT 1011
QY      297 SLN--IADAASLETNMNRTAVTMP 319
Db      1012 TPNDIQADAPSAQSNNEELARVETP 1036

RESULT 10
ID      YD93_METJA          STANDARD:          PRT:          608 AA.
AC      058788:
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein MJ1393.
GN      MJ1393.
OS      Methanococcus jannaschii.
OC      Archaea: Euryarchaeota; Methanococci; Methanococcales;
NCBI_taxid=2190;
[1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8688087;
RA      Bul C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Kertavage A.R., Dougherty B.A., Tomb J.-F., Adams M.A., Reich C.I.,
RA      Overbeek R., Kirkness E.F., Weissman K.G., Merrick J.M., Guyon D.,
RA      Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen A.,
RA      Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RL      Science 273:1058-1073(1996).
CC      -I- SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FUGIDUS AF2028.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U67579; AAB99403.1; -.
DR      PIR: H64473; H64473.
DR      TIGR: MJ1393; -.
FT      TRANSMEM 4 24 POTENTIAL
SQ      SEQUENCE 608 AA: 66768 MW: 010FAFLC29F8C73C CRC64;

Query Match
Best Local Similarity 20.3%; Pred. No. 4.1;
Matches 85; Conservative 68; Mismatches 119; Indels 147; Gaps 25;

QY      1 MAKILF---IFLFFSVLFTRAVSADKP-----GDESTINITEPRDR 41
Db      1 MKRLFLMALMSLIFIGVGY--GDNGPLVAYEYKYNINGNTGGLVSTI----- 52
QY      42 NESSPKHNLNNHTAYSESHLYDRMTFLCISNHT-----LNG-----ACPISE 87
Db      53 -ESTIGYIVINN--TGTTINDLYDVWAVANISNNITPEYVNGTPKGVFISSAPAYT 109
QY      88 N-PSSS-----VSGETNITLQFTEKRLIKRELQIGYKOLLEKSVNCPGLTN-SA 139
Db      110 NLPRANTYIHIPIPLPNNSTVIKFAIDKSTIGVPLINE-----TSDPKITPSEKISNMSV 165
QY      140 HFNCNKAASGASLYLTPAGE-----LKNLPFGGIWDATLKLVRKRYSETGCT 189
Db      166 YLNTSRNYSV-----LPATDTPSVYIMTKYLSNDP-----NNYGS 200
QY      190 YT---INTIKLTGKNIGIOW---LPQFK-----SDARYDLNLRPTGGCT 228

```

```

Db      201 DTWNLNTITGALANGSITIMDGPYFLPGYNDSLTWGVINTKNAITITINI--TGNN 258
QY      229 YIGRSVDMCFYDGTSTNSSSLEIFQDNNPKRSDKFLRLKINDTKRIATY-----LSL 283
Db      259 YTNRTGTLMKY--GEAV-----IFFEFGTKSGTKI-----EGIATYGGYVSA 300
QY      284 ILAGKSLTPNG-----TSLNIADAASLETNMNRTAVTMPDI--SVPLQWPGRLQ 334
Db      301 TREGPFLLAASCKYIWEASNVSKAS--SYFPLTHTITMAVNGSPVILDPENITL 357

RESULT 11
ID      SP21_YEAST          STANDARD:          PRT:          758 AA.
AC      P35209:
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
DE      SP21 protein.
GN      SP21 OR YMR179W OR YMR8010.09.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_taxid=4932;
[1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=94186059; PubMed=8138180;
RX      Natsoulis G., Winston F., Boeke J.D.;
RA      "The SP21 and SP22 genes of Saccharomyces cerevisiae."
RT      Genetics 136:93-105(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / AB972;
RX      PubMed=9169872;
RA      Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA      Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA      Jagels K., Lyne G., Moulé S., Odell C., Pearson D., Rajandream M.A.,
RA      Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;
RT      "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      XIII."
RL      Nature 387:90-93(1997).
CC      -I- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN
CC      YEAST.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: L24436; AAA35078.1; -.
DR      EMBL: Z49808; CAA89912.1; -.
DR      PIR: S47866; S47866.
DR      TRANSEC: T04376; -.
DR      SGD: S0004791; SP21.
DR      GO: GO:0006357; P:regulation of transcription from pol II pro. .; IMP.
FT      DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
FT      DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
SQ      SEQUENCE 758 AA: 84697 MW: 7DB3FCF7EB96705 CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 6;
Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12;

QY      42 NESSPKHNLNNHTAYSESHLYDRMTF-----LCISSNLTNGACPTSENPSSSVS 95
Db      166 NISNKKGRVNNQI-----PEETLEVKLRTFYITNLRTSGNNTNSRISCLQMPSSLT--- 218
QY      96 GETNITLQFTEKRLIKRELQIGYKOLLEKSVNCPGLTNSAHFNCNKAASGASLYL 155

```

Db 219 --PSATLPEPTPKSOSLEKTNQIR-----NSRNARTITIN-----NTNSGT-----257  
 QY 156 YIPAGELKLPFGGIMDPAFLKRYKRRYSEYGYTINTITIKLTKDGNIOIWLPGFKSDA 215  
 Db 258 ---VGRQOTPM---PAPAAVRYQ-----SLPIW-----280  
 QY 216 RVDLNLPTGGGYTIGNSVDMCFY-----DGSTNSSSLLEIFQDNPKSDPKFYLRK 269  
 Db 281 ----NLKPNLANTGFPNRSIAHKIYILADRKTEANQONHONIAEINTIQLNDNTIQRK 336  
 QY 270 INDD--KEIAIYVTSLLAGKSLPTNCT-----SLNI 300  
 Db 337 IDDSVSKRFDFMLNKRKSTKVKSPGIATIAKKPASINI 374

## RESULT 12

NT133\_YEAST STANDARD; PRT; 1157 AA.  
 AC P36161;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 GN Nucleoporin NUP133 (Nuclear pore protein NUP133).  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JUXJR;  
 RX MEDLINE=95112817; Pubmed=7813444;  
 RA Doye V., Wept R., Hurt E.C.;  
 RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+  
 RNA transport and nuclear pore distribution.";  
 RL Yeast J. 13:6062-6075(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9426327; Pubmed=8203164;  
 RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,  
 Ramacia M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;  
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces  
 cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,  
 and six new open reading frames.";  
 RL Yeast 10:231-245(1994).  
 CC -1- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE  
 DISTRIBUTION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X80066; CA56372.1; -;  
 DR EMBL: Z27116; CA81633.1; -;  
 DR EMBL: Z28307; CA82161.1; -;  
 DR PIR: S38160; S38160.  
 DR SGD: S0001790; NUP133.  
 DR GO: GO:0005643; C:nuclear pore; IDA.  
 DR GO: GO:0006406; P:mRNA-nucleus export; IMP.  
 DR GO: GO:0006999; P:nuclear pore organization and biogenesis; IMP.  
 DR Pfam: PF04044; Nup133; 1.  
 DR Nuclear protein; Transport.  
 DR SQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 5.28; Score 98; DB 1; Length 1157;  
 Best Local Similarity 22.18; Pred. NO. 10;  
 Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

QY 48 HNLINHTAY-----SESHLYDRMTF-----LCLSHNTLNGACPTSENP 89  
 Db 96 YGLVNDHKRYIYINNIHSQTQDTPYIVPFSDDNDDEIAVAPRCILFPPATMDSPLALNP 155  
 QY 90 SSSSVSGE-----TNITLQTEKSLKRELQIR-----GYKQLLF 126  
 Db 156 NDDDETGLIILIKSKAIVYEDINSINNLNFKLSEKFS---HELELPINSSGGCKDLM- 211  
 QY 127 SYNC-PSGLTNS-----AHNCKNKAASGASLYIYIPAGELKNLPFG-GIMDATLKR-- 178  
 Db 212 -LNCPEAGIVLSLQDLYPFPAGTUKIMDSHPLODESSQLFSLSYDSSCNFTYILSTPIF 266  
 QY 179 -VKRRYSEY--GYTINTITIKLTKDGNIOIWLPGFKSDARVDLNLPTGGGYTIGNSV 235  
 Db 267 VSLRNGPIILGKGRLYIYI--TNKGIPTW--QLSA-----TNSHPT-----KLI 308  
 QY 236 DMCFYDGYSTNSSSL-----ELRFQDNPKSD--GKFIYRKINDT-KEIAIYVTSLLA 286  
 Db 309 DVNIYEALLESLODLYPFPAGTUKIMDSHPLODESSQLFSLSYDSSCNFTYILSTPIF 368  
 QY 287 GKSLFTPNGSLNADAASLETNNRTIAYTM-PEISVP 324  
 Db 369 DSS-----SNSFTLFTYRLNTEMSITDRFKRPIIP 402

## RESULT 13

CDG2\_PAEMA STANDARD; PRT; 713 AA.  
 AC P31835;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 GN Cyclomalodextrin glucanotransferase precursor (EC 2.4.1.19)  
 DE Cyclomalodextrin glucanotransferase (Cgtase).  
 OS Paenibacillus macerans (Bacillus macerans).  
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
 OX NCBI\_Taxid=44252;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 28-37.  
 RA Sugimoto T., Kubota M., Sakai S.;  
 RT "Polypeptide possessing cyclomalodextrin glucanotransferase  
 activity.";  
 RT Patent number GB2169902, 23-JUL-1986.  
 CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation  
 of a 1,4-alpha-D-glucosidic bond.  
 CC -1- COFACTOR: BINDS TWO CALCIUM IONS.  
 CC -1- SUBUNIT: Monomer.

CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE  
 IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND  
 IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER  
 ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN  
 ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE  
 MALTOOLIGOSACCHARIDE PRODUCED.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC PIR: S26589; ALBSXR.  
 DR HSSP: P43379; ICDG.  
 DR InterPro: IPR006589; Alp\_amy1\_cat\_sub.  
 DR InterPro: IPR006048; Alpha\_amy1\_C.  
 DR InterPro: IPR006047; Alpha\_amy1\_cat.  
 DR InterPro: IPR002044; CBD\_4.  
 DR InterPro: IPR006046; Glyco\_hydro\_13.  
 DR InterPro: IPR002909; IPT\_TIG.  
 DR Pfam: PF00128; alpha-amy1ase; 1.  
 DR Pfam: PF02806; alpha-amy1ase\_C; 1.  
 DR Pfam: PF00686; CBM\_20; 1.  
 DR Pfam: PF01833; TIG; 1.  
 DR PRINTS: PR00110; ALPHAMYLASE.  
 DR PRODOM: PD001568; CBD\_4; 1.  
 DR SMART: SM00642; Amyy; 1.  
 DR SMART: SM00632; Amyy\_C; 1.  
 DR Transferrase; Glycosyltransferase; Calcium; Signal.  
 FT SIGNAL 1 27







```

QY 135 TINSAPHCNKNASGASLYIIPAGELNLPFGIMDAFLKIRVRRYSEYGYTTINI 194
Db 347 VLKTYGYY---DKTTGTN-YAFMKFNNLKPIONMTYTKATLKTYVAHSY---YGT----- 393
QY 195 TIKLFDKGNIOIWLPOFKS---DARVDNLNRPYGGTYIGRNSVDMCFYDGY-----S 244
Db 394 --KATG-----LMLDTVNSNYDNNAKYTWNTKFPASKN--IGKADYHKGMASHDYTAAVKS 444
QY 245 TNSSSLEIRFQDNNPKSDGKFFYLKRI-----NDDTKEIAVTLISLLAGKSLTPT----- 293
Db 445 WNSGANYGFK-LHTNGNGKEYWKLISSANSANKRPIEYTYTIP-----KGNTPTIKAY 498
QY 294 -NCTSLINADASLETNNMRI 313
Db 499 HNGDSTGYFD-----ISMKKV 514

```

Search completed: August 19, 2003, 09:46:26  
 Elapsed time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:43:34 ; Search time 99 Seconds

(without alignments)  
940,980 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886  
Sequence: 1 MNKILFTLFFSSVLFTFA.....EAGQYMGINVTFTPSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_23:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacteriophage:\*  
18: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2	Q93G67 escherichia
2	949.5	50.3	363	2	Q47115 escherichia
3	930.5	49.3	364	2	Q47119 escherichia
4	381	20.2	387	2	Q8KQ77 burkholderi
5	371	19.7	387	2	Q8KQ10 burkholderi
6	275.5	14.6	359	16	Q9XDS1 salmonella
7	127	6.7	1881	16	Q8RKS2 fuscobacteri
8	115.5	6.1	700	2	Q8VUS0 chlamydia p
9	115.5	6.1	847	2	P71132 chlamydia phi
10	114	6.0	865	15	Q8G7H7 human immun
11	113	6.0	2646	5	Q81521 plasmodium
12	111.5	5.9	743	16	Q8XNM2 clostridium
13	111	5.9	2768	16	Q8E9G6 shewanella
14	110.5	5.9	589	16	Q8F994 leptospira
15	110.5	5.9	1052	17	Q8Q0S0 methanosa
16	110	5.8	265	2	Q8KGS5 rhizobium 1

17	110	5.8	846	2	P71133 chlamydia phi
18	109.5	5.8	2454	5	Q8T2G3 dictyosteli
19	109	5.8	807	5	Q8T7V5 drosophila
20	108	5.7	872	2	Q9RLA0 rickettsia
21	108	5.7	2529	16	Q25579 helicobacte
22	107	5.7	304	16	Q8XAX2 escherichia
23	106.5	5.6	1012	16	Q8PER7 xanthomonas
24	106.5	5.6	1065	16	Q8EXX0 leptospira
25	106	5.6	1937	5	Q8IE94 plasmodium
26	105.5	5.6	430	3	Q96VM2 blumeria gr
27	105.5	5.6	843	16	Q8YKQ8 anabaena sp
28	105	5.6	398	16	Q8Y7I7 listeria mo
29	104.5	5.5	1441	16	Q9CFL1 lactococcus
30	104	5.5	304	16	Q8CWZ7 escherichia
31	104	5.5	803	10	Q9AUM5 oryza sativ
32	103.5	5.5	691	16	Q9RZS7 delnoccocus
33	103.5	5.5	1269	17	Q97Z06 sulfolobus
34	103.5	5.5	2399	16	Q9ZKS9 helicobacte
35	103	5.5	868	17	Q8PM5 methanosa
36	103	5.5	1754	16	Q8RD81 thermococ
37	102.5	5.4	1736	5	Q95PH7 dictyosteli
38	102.5	5.4	2747	2	Q91800 aeromonas s
39	102	5.4	376	5	Q19229 caenorhabd
40	102	5.4	692	17	Q97YM6 sulfolobus
41	101	5.4	881	16	Q8EW44 mycoplasma
42	101	5.4	1238	16	Q8ZJ21 yersinia pe
43	101	5.4	1267	16	Q8CLQ7 yersinia pe
44	101	5.4	2082	5	Q812R1 plasmodium
45	100.5	5.3	914	16	Q8EX32 mycoplasma

# ALIGNMENTS

## RESULT 1

ID Q93G67 PRELIMINARY; PRT: 361 AA.  
AC Q93G67;  
DT 01-DEC-2001 (TREMBLER, 19, Created)  
DT 01-DEC-2001 (TREMBLER, 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLER, 19, Last annotation update)  
DE Minor pilin protein CsaE.  
GN CsaE.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Bacteria; Proteobacteria; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=E11881A;  
RA Altboum Z.D., Levine M.M., Galen J.E., Barry E.M.;  
RT "Isolation and characterization of EPEC CS4 fimbriae encoding genes,  
RT and their expression in Shigella flexneri 2a quana strain CVD 1204.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF296132; AAK97137.1; ZIE174ABDB0EB353 CRC64;  
SQ SEQUENCE 361 AA; 40102 MW; 2E1E74ABDB0EB353 CRC64;

Query Match 100.0%; Score 1886; DB 2; Length 361;  
Best Local Similarity 100.0%; Pred. No. 7.1e-137;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNKLFTLFFSSVLFTFAVSADKIPGDESTINIFGPRDRNESSPRHNILNHNITAYSE	60
DB	1	MNKLFTLFFSSVLFTFAVSADKIPGDESTINIFGPRDRNESSPRHNILNHNITAYSE	60
QY	61	SHITLDRMTFLCLSSNHTLNGACPTSPNPSSSVSGTNTITLOFTKRSIKRELQKGY	120
DB	61	SHITLDRMTFLCLSSNHTLNGACPTSPNPSSSVSGTNTITLOFTKRSIKRELQKGY	120
QY	121	KOLLFKSVNCPGSLTNSAHFNCKNKAASASLYLYIPAGELKNLPPGGITMDATLKLRYK	180
DB	121	KOLLFKSVNCPGSLTNSAHFNCKNKAASASLYLYIPAGELKNLPPGGITMDATLKLRYK	180

Oy		181	RYSSETYGTATNTIKTLTKDKNIOIWLPOFSDARVDLNLPTGGCTIGRNSVMCEY	240
Dd		181	RRYSTYTGYTNTIKTLTKDKNIOIWLPOFSDARVDLNLPTGGCTIGRNSVMCEY	240
Oy		241	DGYSTNSSSLEIRFODNNPKSGDKFYLRKINDTKEIAVTTLSLLAGKSLPPTNGTSLNI	300
Dd		241	DGYSTNSSSLEIRFODNNPKSGDKFEYLKRINDTKEIAVTTLSLLAGKSLPPTNGTSLNI	300
Oy		301	AADASLETNMNRITAVTMPEISVPILCMGRLOLDAKVENPAGQVMGNINTEFPSSOT	360
Dd		301	AADASLETNMNRITAVTMPEISVPILCMGRLOLDAKVENPAGQVMGNINTEFPSSOT	360
Oy		361	L 361	
Dd		361	L 361	

RESULT 2

047115 PRELIMINARY; PRT; 363 AA.

047115;

01-NOV-1996 (TREMBLrel. 01, Created)

01-NOV-1996 (TREMBLrel. 01, Last sequence update)

01-MAR-2003 (TREMBLrel. 23, Last annotation update)

COO precursor.

COO.

Escherichia coli.

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

NCHI\_TaxID=562;

[1]

SEQUENCE FROM N.A.

MEDLINE-94344028; PubMed-7915003;

Friedrich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;

"Cooc and Cooc are required for assembly of CSI pilin."

Mol. Microbiol. 12:387-401(1994).

EMBL; X76908; CA54320.1; "

Interpro: IPR001412; tRNA-synt.1.

DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; 1.

KM Signal.

FT SIGNAL.

FT CHAIN

SQ SEQUENCE 363 AA; 40139 MW; AA9488AADDDEDF72 CRC64;

Query Match 50.3%; Score 949.5; DB 2; Length 363;

Best Local Similarity 53.1%; Pred. No. 6, 2e-65;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11

3 KILEFTLFESSVLETFPAVSADKIIPGDE--SITNIF-GPR-DRNESPKNHNLNNHTAY 58

2 KIIFFL-----SIIPSAVVASGRYPETVTGNLTSEFOAPRLDRSVQSVPINFTNHVAY 57

59 SESHTLVRMTFLCLASHNTLNGACPTSENPESSSVS-GETNITLOFTFKRSLIKRELOI 117

58 SLSHSLYRIAYFLCSSLSSNPVAGAPT-I--GTSVOYGTTTITLOFTEKRLIKRINKIMVL 114

118 KGYYQLLEFSYNCPESG--LFLNSAHFNCKNKNA-ASGASLYLYTPAGELNLPFGGITMAT 174

115 AGNKKPIWENQSCDPESNLMLVANSKSMSCGAHGHNANGTLLNLXITPAGEIKKLFPQGIWMT 174

Oy 175 LKLAVKRRYSET----YGTYYTNITIKTLTKDKNIOIWLPOFSDARVDLNLPTGGCTIY 230

Dd 175 LILRLS-RYGEVSSHRYGNIVTYNITVDTLTKDNIOIWLPGFHSNRPVDLNLPTIGNYKES 233

Oy 231 GRNSVDMCFYGYGSYTNSSSLEIRFODNNPKSGDKFYLRKINDTKEIAVTTLSLLAGKSL 290

Dd 234 GSNSLDMCFYGYGSYTNSSSLEIRFODNNPKSGDKFYLRKINDTKEIAVTTLSLLAGKSL 292

Oy 291 TPTNGTSLNIADAASLETNMNRITAVTMPEISVPILCMGRLOLDAKVENPAGQVMGNI 350

Dd 293 YPVNGQSFINDSSVLETFNMNRITAVTMPEEVNVPIVLCMPARILLNADVNA PDAGQYSGOI 352

Oy 351 NVTFPPSQTL 361

[illegible]

Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;

DR EMBL; AJ242964; CAB51577.1; -.

KW Hypothetical protein; complete proteome.

SQ SEQUENCE 359 AA; 39741 MW; BB88A4E9A62052CE CRC64;  
 Query Match 14.6%; Score 275.5; DB 16; Length 359;  
 Best Local Similarity 29.7%; Pred. No. 3.6e-13;  
 Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;  
 QY 70 FCLSHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLIKRELQINGYKOLLFSYV 129  
 Db 79 WVCRSNRNENEGACEETHLWMWYAFGAYSKIRLRFREQISHAEITL-----ILLGSR 131  
 QY 130 --CPGSLTNSAHFNCKNAAS---GASLYIIPAGELKNLPFGGIMDTLKL-RKRRY 183  
 Db 132 DACYGV-----INMNAACQWGRSLKRLPSEELAKIPSGWKATLVLDYLOMG 183  
 QY 184 SETYGYTITITIKLTD-KGNIQIWLPOF-KSDAVDLNLRPTGGCTYIGRNSVDMCFY 240  
 Db 184 DDPGLGSTDITLANTDHEAENAAITFPQFGATRPVDDLNRHMASQMSGRANLDMCLY 243  
 QY 241 DGSTNSSLEIRFODNNPKSDGKEYLRKINDTKEIATYLSLLAGKSLPTNGTSLNI 300  
 Db 244 DG-GVKARSLQMKIEGSKNGSGTGFOVIRKSDSADT--IDYAVSMNYGSRISIPYRGVEFSL 300  
 QY 301 ADAASLETMMNRITAVTMEISVPYLCWPGRIQLOAK---VENPEAGOVNMINVTFTS 357  
 Db 301 DWVKAATR-----PYVLPGRQAVACVPLPLTTTPPNIRKRGSEYQTLTPTMLNG 355  
 QY 358 SQT 360  
 Db 356 TOT 358  
 RESULT 7  
 Q8RGK2 PRELIMINARY; PRT; 1881 AA.  
 AC Q8RGK2;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hemolysin.  
 GN FN0291.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
 OC Fusobacterium  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25586;  
 RA MEDLINE-21886394; PubMed-11889109;  
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyripides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586."  
 RT J. Bacteriol. 184:2005-2018(2002).  
 DR EMBL; AE010541; AAL94497.1; -.  
 DR InterPro; IPR001452; SH3.  
 DR InterPro; IPR000508; SigPase.  
 DR Pfam; PF00018; SH3; 1.  
 DR PROSITE; PS00761; SPASE\_1\_3; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 1881 AA; 204375 MW; D635156A4EFA4877 CRC64;  
 Query Match 6.7%; Score 127; DB 16; Length 1881;  
 Best Local Similarity 23.1%; Pred. No. 0.78;  
 Matches 86; Conservative 45; Mismatches 120; Indels 122; Gaps 19;  
 QY 1 MNKILFTLFSSVLETFV-----SADKIPGDESITNIFGPRDRNRSSEKRN- 49  
 Db 5 LKKLTAIFMLFHIISLADGIVDNSAKSKNLQVDRKANGVPLVNIAP-DNNGTS--HNW 61  
 QY 50 -----ILNHNHTAYSESHLYDNTFLCLSHNTLNGACTSENPSSSSVSG-- 96

Db 62 YKDYVNDGRGALNN-----SDDLNSQAGGLI 89  
 QY 97 ENRITLQFEKRSLLIKREL-----QIKYKOLLKSVNC-----PSGLTNSAHF--NKN 144  
 Db 90 YGNPNLQNSKEASTIINEVGVNKSRIEYQIYKRAYIILANPGLIYNAGFTNGN 149  
 QY 145 KNAASGASLYIYIP-----AGELKNLPFGGIMDTLKLRRYRSETYGYTITITIK 197  
 Db 150 VFTTGRGNLNLNPERGMIEING-KGLDLRINNAELIARAEISAPYIGGEVYNLKG 207  
 QY 198 LTKGNIQIWLPOFSDARVDLNRPTGGCTYIGRNSVDMCFYDGSTNSSLEIRFO-- 255  
 Db 208 NQGSN---KPEYALDAR-----ALGSIYAGRINI-----IYVEDGVKTOAP 248  
 QY 256 -----DNNPKSDGKRYLR-----KINDTKEIATYLSLLAGKSLPTNGTSLNI 301  
 Db 249 MYATKGDVYISSKRGVYLKDTQAKRDIKISTETEIG--SKLAEINAIKSGKTSN-- 303  
 QY 302 DAASLETMMNRIT 314  
 Db 304 -SGQIRAN-NNTT 314

RESULT 8  
 Q8VU50 PRELIMINARY; PRT; 700 AA.  
 ID Q8VU50;  
 AC Q8VU50;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Putative polymorphic membrane protein (Fragment).  
 OS Chlamydia psittaci (Chlamydia psittaci).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83554;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-POS;  
 RA Laroucau K., Souriau A., Rodolakis A.;  
 RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in  
 RT serotype-1 Chlamydia psittaci strains."  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBD databases.  
 DR EMBL; AF243415; AAL36959.1;  
 DR InterPro; IPR003368; Chlamydia\_pmp.  
 DR Pfam; PF02415; DUF145; 1.  
 DR TIGRPFAM; TIGR01376; POMP\_repeat; 3.  
 FT NON-TER 700  
 SQ SEQUENCE 700 AA; 74200 MW; 2C38E504AE48D1D1 CRC64;  
 Query Match 6.1%; Score 115.5; DB 2; Length 700;  
 Best Local Similarity 18.3%; Pred. No. 1.7;  
 Matches 73; Conservative 63; Mismatches 124; Indels 139; Gaps 18;  
 QY 42 NESSPEKHNINHNHTA-----YSESHLYDNTFLCLSHNTLNGAC----- 83  
 Db 27 NETLISSDSYINGVNSDEFYKETTSGAIYITBGNVCISYAGKDSPLNCSCTETENIS 86  
 QY 84 -----PTSENPSSSVSGETNITLQFTKRSLLIKRELQINGYKOLLFSYV 129  
 Db 87 FIGNGYTLCFDNTTASNPALINVSQD-----QKTLNVGSFS--LFSCH 130  
 QY 130 CPSSGLT-----INSAHF--NCKM-----NAASGASLYIYIP-----A 159  
 Db 131 CPPTGTGYGAIQTKGVSTFGNNKLIFDNNCSTGEGAIKCATGSNAELKTEGNSYVVS 190  
 QY 160 GELKNLPFGGIMDTLKLRRYRSETYGYTITITIKLTDGNGNQIWLPOFGRSDARVDL 219  
 Db 191 GNSSQKGGAIY--TKLITITADGPTLSENNNSVAS--SPKGG-AICLDOTSSSCSLTA 244  
 QY 220 NT-----RPTGGCTYIGRNSVDM-----CFYDGSTNSSSLE 251  
 Db 245 NIGDITFDGNKYIKINGSSSYVKNRAIDLGSQGRFTKLNAKRGFIFHDPILANNGSRE 304

```

QY      252  IRFQDNNPSPD-----GK--FYLRKINDDTKEALAYL-----STLLA 286
Db      305  IEL--NKTESDPTTYTGKIVFSGEKISDEBKEYPDMKLKSYFKQPLKIGAGSVLKDQSVTL 362
QY      287  GKSLPTNGTSLNIDAPASLETNNMNRITAYMPELSVEY 325
Db      363  AKKITOTKGSIVMDGCTIQTLPSSGGETITLLTNLDINI 401

```

## RESULT 9

	PRT:	847 AA.
--	------	---------

```

AC P71132; PRELIMINARY;
DT 01-FEB-1997 (TrEMBLrel_02, Created)
DT 01-FEB-1997 (TrEMBLrel_02, Last sequence update)
DT 01-MAR-2003 (TrEMBLeRel_23, Last annotation update)
DE POMp91A.
OS Chlamydomophila abortus.
OC Bacteriota; Chlamydiota; Chlamydiales; Chlamydiaceae; Chlamydophilia.
OX NCBI_TaxID=83555;
[1]
SEQUENCE FROM N.A.
STRAIN=S26/3;
MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins
RL of the ovine abortion subtype of Chlamydia psittaci."
RN FWMS Microbiol. Lett. 142:277-281(1996).
[2]
SEQUENCE FROM N.A.
STRAIN=S26/3;
MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT RT "Molecular cloning and characterization of the genes coding for the
RL highly immunogenic cluster of 90-kilodalton envelope proteins from the
Chlamydia psittaci subtype that causes abortion in sheep.";
INfect. Immun. 66:1317-1324(1998).
DB EMBL; U65942; AACI5921.1; -
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR003368; Chlamydia_Pmw.
PFam: PF02415; DNP145; 1.
DR TIGRfams; TIGR01414; autotrans_barl; 1.
DR TIGRfams; TIGR01376; POWP_repeat; 3.
SQ SEQUENCE 847 aa; 90694 Mw; 754C958B7F11179E CRC64;
```

Query Match	6.1%;	Score 115.5;	DB 2;	Length 847;
Best Local Similarity	21.6%;	Pred. No. 2.1;		
Matches	94;	Conservative	58;	Mismatches 153;
				Indels 131;
				Gaps 21;

Db 9 LISSLIIVANSIYSEE-EDOKTILTAHSYNGNTNSEP-----FNPILSTSNIGTIYVCTG 62  
 QY 70 FCLIS-----SHNTLGNACPT-----SENSSSSVSGEITITL 102  
 Db 63 NCIIVAGLDGSGLSSSCFPTACNLSPFLGNGYTLCPDNITPQSSHHGALISVGS-TNKT- 120  
 QY 103 QTEFRSLIKRLQIKGYKOLLFKSVNCPSLTINSA-----HFCKNK 145  
 Db 121 -----LDISGFS--LFSCAVCPGATGYGAIKAVGNTTIKDNSSLVFHKKCS 166  
 QY 146 NAAAGASILYLIYPAGELK-----MLPF-----GGIMPATLTLRVKRRYSSETGYTYI 192  
 Db 167 GEGGAIQCKASSSEAELEKIENNOMLVFENASSSSGGAIVAD-KLTIIVSGGPTLFESNSV 225  
 QY 193 NIT-----IKLTGDNQIQLMPQKSPARVDLN--LRPGGGGYIIRNSVDM----- 237  
 Db 226 SASPSRGKALICIKDGG-ECSLTADLDGITPDGKKIIKTNGSGPTVIRNSIDLGSGCKT 264  
 QY 238 -----CYDGS--TNSSSEIIRQDNNPXSODK--FYLRKIMDTKELAYTL-- 281  
 Db 285 KINAEGRGIEIFYDITGGGDELININKQDVPVDT-KTIVFSGRRSLDEKKKVAANLKSD 343

```

QY      282  -----SLLG-----KSLTPNGSLNDAASLET--NNNRIRAVMPETISV 32
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      344  FKQPLKIGSGSLIKDGVLTETKSTQTBGATVADLTQLTSSGGFTITLINIDINV 403
          ||| | | | | | | | | | | | | | | | | | | | | | | | |
QY      324  PVLICMPGRLLDRAVE 339
          ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      404  ASLGGGGVADPAKVE 419
          ||| | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 10

ID	0807H7;	PRELIMINARY;	PRT;	865 AA.
AC	0807H7;			
DT	01-JUN-2002 (TrEMBLrel. 21, Created)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Envelope glycoprotein (Fragment).			
GN	ENV.			
OS	Human immunodeficiency virus 1.			
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.			
OX	NCBI_TaxID=11676;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=99CMA304.			
RX	MEDLINE=21849375; PubMed=11860674;			
RA	Yamaguchi J., Vallari A.S., Swanson P., Bodelle P., Kapte L.,			
RA	Manosp C., Zekeng L., Guttler L.G., Devare S.G., Brennan C.A.;			
RT	"Evaluation of HIV Type 1 Group O Isolates: Identification of Five			
RT	Phylogenetic Clusters.";			
RL	AIDS Res. Hum. Retroviruses 18:269-282(2002).			
RL	EMBL; AF383244; AAL98866.1; -			
DR	InterPro: IPR000328; ENV_GP41.			
DR	InterPro: IPR000777; GP130.			
DR	Pfam: PF00516; GP120; 1.			
DR	Pfam: PF00517; GP41; 1.			
KW	AIDS; Coat protein; Glycoprotein; Polypeptide; Transmembrane.			
FT	NON_IDR			
FT	SEQUENCE 865 AA; 97118 MW; 996B83862AA51ACA CRC64;			

Query Match	6.0%;	Score 114;	DB 15;	Length 865;
Best Local Similarity	21.5%;	Pred. No. 2.9;		
Matches 93;	Conservative 56;	Mismatches 150;	Indels 134;	Gaps 22

```

QY      21 V$ADKJPGDBESTJFIPEDRNESSPKNIILNNH-----TTA$E$H$TYD$RNTFL 71
D      75 V$PDPPIRYPPYPLHN-----IDK$NI$EN$M$K$Q$G$D$IIDL$M$Q$SLK$PCV$KMTFL 125
QY      72 C$SH$NT-LNG$ACPT$ENP$SS$VS-----G$T$NITL$PTE$RS$L$IK$REL$Q$K$Y$K$D$LFK 136
D      126 C$Q$M$CT$SL$N-----D$K$SN$T$S$P$E$M$K$CE$N$T$Y$V$K$D$K$E$K$Q$K$AL$F$V$S$D$M$K$D$K 182
QY      127 S$V$N$C$B$G$L$T$N$A$H$N$C$K$N$A$S$G$-----S$L$Y$Y$P$A$G$-----E$L$K$N$D$P$G$C$--- 169
D      183 N$T$N$T$M$Y$LL$-----N$C$N$S$T$T$S$Q$C$P$Y$S$E$P$I$P$H$C$A$G$A$Y$A$F$K$C$N$N$IE$N$G$T$C 238
QY      170 -----I$M$D$A$L$-----K$L$R$V-K$R$K$S$E$Y$G-T$Y$T$N$T$IK$Y$D$K--G 202
D      239 K$N$T$V$V$T$C$H$G$I$K$P$V$S$T$Q$L$I$N$G$L$T$K$G$K$R$I$M$T$K$N$D$V$N$T$Y$V$N$T$S$T$K$IT$C$K$R$P$G 298
QY      203 N$T$Q$M$L$P$Q$K$S$A$R$D$L$M$L$P$T$G$G$T$Y$G$R$N$S$V$D$M$C$F$D$G$S$T$N$S$S$S$L$E$I$R$F$O$D$N$N$K$S$D 262
D      299 N$M$T$V$O$V$S$I$G$P$A$M$Y$S$M$D$R-----G$T$E$R$N$T$R$V$A$Y$C$E$-----N$S$T$W$E$-----R$T$L$K$O$T 344
QY      263 G$K$F$L$K$I$D$D$P$K$E$I$A$Y$T$L$S$LL$G$-----K$S$L$-----T$P$T$Q$S$L 298
D      345 A$E$R$F$E$L$V$N-N$T$K$K$V$D$M$T$F$N$S$G$G$D$E$V$A$N$L$H$N$C$G$E$F$F$Y$C$M$T$S$S$L$F$Y$T$T$S$C$M$G$S$T$C 403
QY      299 N$I$A$D$A$S$L$E$T$N$M$N$R$I$-----T$A$V$M$P$E$I$S$V$P$V$L$C$M$P$G$R$L$D$A$K$E$V$N$P$E$A 343
D      404 N$T$K$S$N-S$T$N$T$R$I$P$C$R$L$Q$V$Y$K$W$I$O$G$G$G$L$A$P$P$I$R$G$N$T$C$M$S$N$T$I$M$L$Q$M$Q$P-- 460
QY      344 G$O$Y$M$G$N$V$N$T$P$P 356

```



DB 461 ---WNHSNATFRP 470

# RESULT 11

081521 PRELIMINARY; PRT; 2646 AA.

01-MAR-2003 (Tremblrel. 23, Created)

01-MAR-2003 (Tremblrel. 23, Last sequence update)

01-MAR-2003 (Tremblrel. 23, Last annotation update)

Erythrocyte membrane protein 1 (PFEMP1).

PEL1950W.

Plasmodium falciparum (Isolate 3D7).

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

NCBI\_TaxID=36329;

SEQUENCE FROM N.A.

STRAIN=3D7;

MEDLINE=2225705; PubMed=12368864;

Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Paul A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S., Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.

"Genome sequence of the human malaria parasite Plasmodium falciparum."

RL Nature 419:498-511 (2002).

EMBL: AE014850; AA036476.1;

SEQUENCE 2646 AA; 300284 MW; C91D257F3DEF8717 CRC64;

Query Match

Best Local Similarity 24.8%; Score 113; DB 5; Length 2646;

Matches 77; Conservative 37; Mismatches 128; Indels 68; Gaps 15;

DB 28 GDESTNIFGPRDRNESSPKHNLNHTAYSESHLYDRMTFLCLSSHTNLGACPTS- 86

DB 695 GDS--TAIGIKKLEKKKKKKGAN-----ISEKILDEFLNHELDKATCKCPKRF 747

DB 87 EHPSSSVSGENITITQTEKRSLLKRELQIKGYKQLLFKSVNCPGGLTNSHFNCNN 146

DB 748 KNPCCSDTSGDSN--KQYAVANTVAQIIQGRKQOL-----HNGSSN 789

DB 147 AASGASLYLYIPAGELKNLPFGIMDATLKLKRYSEYTYTITIKLTKDKNI-- 204

DB 790 ALKGNQNAKKNKGRPN-----PLTACQITKHSNGK---DSNNPCNNKGNRLK 838

DB 205 --QIWLPOKSD-ARYDL-----NLPTGGGTYYIGRNSVDMCFYDG--YSTNS 247

DB 839 ICGVW--SIKNDTSYDYVMPRRQHMCTSNLEKLYASVIGSNVNDKFLVEVLHAKS 896

DB 248 SLELRFQNNKPSCKEFLKIKINDT-KEIAYT--LSLLAGKSLPTTNGSLINIAA 303

DB 897 EAEFLIKKRYKNGKNGKRLKRDQATTCRAIRYSPADIGDIIRGKDLMDNN-----DA 950

DB 304 ASLETMMNRI 313

DB 951 KSLQTNLRAI 960

RESULT 12

08XNM2 PRELIMINARY; PRT; 743 AA.

01-MAR-2002 (Tremblrel. 20, Created)

01-MAR-2002 (Tremblrel. 20, Last sequence update)

01-MAR-2003 (Tremblrel. 23, Last annotation update)

Hypothetical protein CPE0220.

CPE0220.

Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

NCBI\_TaxID=1502;

RP SEQUENCE FROM N.A.

RC STRAIN=13 / Type A;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kohara S., Hayashi H.

"Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater."

Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).

DR EMBL: AP003185; BAB79926.1;

DR InterPro: IPR001899; Gram\_pos\_anchor.

DR SMART: SM00725; NEAT\_4.

DR TIGR01167; LPXNG\_anchor; 1.

KW Hypothetical protein; Complete proteome.

SEQUENCE 743 AA; 81149 MW; A7C451E9C5A5E545 CRC64;

Query Match

Best Local Similarity 18.8%; Score 111.5; DB 16; Length 743;

Matches 64; Conservative 66; Mismatches 115; Indels 95; Gaps 13;

DB 7 ITLFFSSVLTFFNVSADKIPGDESTITFGPRDRNES---SPKHNLTNNHTAYSES 61

DB 88 LMTVYFNSLYFGFMNNEIVSAGGEALKEENKDDSTIREVPSPTKRYIGLFTMMGRK 147

DB 62 HTLYDEMTFLCLSSHTLN--GACPTSEPPSSSV-----SG-----ETNITLQ 103

DB 148 VELF-----LYNDMNTVLLDEAPLNNAKDISTVQGAIDLSGIVGTCKEDSNLKYE 201

DB 104 FTEKRSLLK-RELQIKGYKQLLFKSVNCPGGLTNSHFNCNNKMAASGASLYLYIPAGEL 162

DB 202 ISGDTSEFVADGKAIEIGVPIITYK-----VTDSSGQD----- 234

DB 163 KNPFGIMDATLKLKRYSEYTYTITIKLTKDKNI---ITLPOKSDARD 218

DB 235 -----EKTVMYVNNKKTTLGDSSTYLLKNTVVOYVGGNMETSNAKRYLSDSRID 285

DB 219 LNLPTGGGTYYIGRNSVDMCFYDGYSTNSLEIRFODNPNKPSDKGYFLKINDPTKEIA 278

DB 286 IS-----NKNNTVITLF-----NSELTAPLKNNTVYDGEKVAEYKNNKRTIK 329

DB 279 YTL-----SLLAGKSLPTTNGSLINIAASLETN 309

DB 330 FNILPDLNSDIYVTLVSMGKREVSFKTILNVTAKLEDN 369

# RESULT 13

08E9G6 PRELIMINARY; PRT; 2768 AA.

01-MAR-2003 (Tremblrel. 23, Created)

01-MAR-2003 (Tremblrel. 23, Last sequence update)

01-MAR-2003 (Tremblrel. 23, Last annotation update)

RTX toxin, putative.

SO4317.

Shewanella oneidensis.

Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

Alteromonadaceae; Shewanella.

NCBI\_TaxID=70863;

SEQUENCE FROM N.A.

STRAIN=MR-1;

MEDLINE=22297686; PubMed=12368813;

Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S., Beyer R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamthyan J., Weidman J., Imprial M., Lee K., Berry K., Lee C., Mueller J., Khouli H., Gill J., Uterback T.R., McDonald L.A.,

RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RL *Shewanella oneidensis*.";  
 DR EMBL: AEO1118-1123(2002).  
 DR EMBL: AEO15864; AAM57286.1; -  
 DR TIGR: SO43177; -  
 KW Complete proteome.  
 SQ SEQUENCE 2768 AA; 285094 MW; 55FCE30E3E09C1BE CRC64;

Query Match 5.9%; Score 111; DB 16; Length 2768;  
 Best Local Similarity 19.1%; Pred. No. 22;  
 Matches 81; Conservative 68; Mismatches 160; Indels 116; Gaps 20;

QY 13 SSVLTF--AVSADKIPGDEITNIGPRD-----RNSSPKHILNNHTAYSES 61  
 DB 1952 SSALLTINVTVNANNAKAVODOKN--GPDIAVIGVLENDGKNNVHFHTANSAIQ 2009  
 QY 62 HTLYDRNTF---LCSSHNTL-----NGACPTSENPSSSVSEETNITLOFTEKRS 109  
 DB 2010 HTAGSKITTLAEGELTLNADGSYTFETPAADNNGQPVITTYTNGATSTLTIVTPVDDPT 2069  
 DB 110 LIRKELQ-----IKGYKQLFKSVNCPGSLTINSAPNCKNAAGASLYLTPAGEL 162  
 DB 2070 VTRPESITIAEDALAKG--NVLENDIVDNTLSTVS--FOVNGVTVNAGNWTOLPGG-- 2123  
 QY 163 KNLPGGIMDATLKLKRVKRRYS-----ETVGYTINITIK----- 197  
 DB 2124 -----TLQLKTNGEISFPDKHWSGSLPVITTYTNTGATSTLTIVQAVADAPN 2172  
 QY 198 LTKGNQIQLPQKSKARD-----LNLRPT-----GGGYTIGNSVDMC 238  
 DB 2173 LTNGYSVAALINE--DAKFGSGWDGVKANDIKGLNTIGWHTSNRGQVEIGESV--- 2228  
 QY 239 FYDYSTNSSSLERFODNPP-----KSD-GKPY-----LRKINDTKEIAYTLSTLLA 286  
 DB 2229 YVSGSGSNKVMETEEFNSGDKITLYTDIQADAGRTELGFDIAANSVSVSGLTIKILPL 2288  
 QY 287 GKSITPTNGTSLNIAADASLETNNMRTAVTAPETISVPLCWPRQLDIAKVENPEA--GQ 345  
 DB 2289 DAQGNPLNAQAITLYDEDPNANMLRQOKITLP-----INOSKRYLRQEGDNGDSYGA 2342  
 QY 346 YMGNI 350  
 DB 2343 LLDNL 2347

## RESULT 14

QY 08F994 PRELIMINARY; PRT; 589 AA.  
 DB 08F994; 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DE Outer membrane protein ompA family.  
 GN LA0301.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / serogroup Icterohaemorrhagiae / serovar 1a1;  
 RL Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO11218; AAM47500.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 589 AA; 67901 MW; 26A00B05B7E73DC CRC64;

Query Match 5.9%; Score 110.5; DB 16; Length 589;  
 Best Local Similarity 21.5%; Pred. No. 3.2;  
 Matches 71; Conservative 50; Mismatches 109; Indels 101; Gaps 16;

QY 40 DRNESSPKHILNNHTAYSESHT-----LYDRMTFLCLSSHNTLNGACPTSENPSSSS 93

DB 132 ENNE-----NLSEKTIILDSHSPSSSEIGKTVLE-----LNGDRNSSRINISOK 178-  
 QY 94 VSGETN--ITLOPEKSLIKRELQIKGYKQLFKSVNCPGSLTINSAPNCKNAAGAS 151  
 DB 179 VFGESSDYFTGEIFLDSGKY-----VKSY---TWKLADVBSVLIMDGTDES--NKQLSNGI 229  
 QY 152 SLVYLIPAGELKMLPFGIMDATLKLKV-----KRRSEYTG-----YTINITIK 197  
 DB 230 YTYRLISGDKVKNNSISVSLDITRNEMIGVDMFSDSKITSYSGSLKNSARFTFYSPK 289  
 QY 198 L-IDKGNQIQLPQKSDARYDLNLPTG-----GGTYIGRNSVDM 237  
 DB 290 LKSDSYELIEVFOKNGNEKTV-YRLRETGEPEWKNLKNQIGDFVSAGTYFCRLTVH- 347  
 QY 238 CFYDGYSTNSSSLERFODNPPKSDGKFYLRKINDPKEIAYTLSTLLAGSLPTNGTS 297  
 DB 348 SRYRYQSIPTSSFT-----SKF-SFNLDLSLTKERSPNDGK 385  
 QY 298 LNIADAASLETNNMRTAVTAPETISVPLCW 328  
 DB 386 -----NDLIRIYLSHOGIPLOSW 403

## RESULT 15

QY 08Q050 PRELIMINARY; PRT; 1052 AA.  
 DB 08Q050; 01-OCT-2002 (TREMblrel. 22, Created)  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Conserved protein.  
 GN MM0066.  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartesch T., Merl R., Schmitz R.A.,  
 RA Martinez-Artas R., Henne A., Wlezer A., Baumeister S., Jacobl C.,  
 RA Brueggemann H., Llenard T., Christmann A., Boemcke M., Steckel S.,  
 RA Bhattacharyya A., Lykilds A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazel: evidence for lateral gene  
 transfer between Bacteria and Archaea.";  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL: AEO13226; AAM29762.1; -  
 DR InterPro: IPR000601; PKD\_domain.  
 DR Pfam: PF00801; PKD; 2.  
 DR PROSITE: PSS0093; PKD; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 1052 AA; 112882 MW; DB84642718DF2121 CRC64;

Query Match 5.9%; Score 110.5; DB 17; Length 1052;  
 Best Local Similarity 22.9%; Pred. No. 6.8;  
 Matches 70; Conservative 52; Mismatches 115; Indels 69; Gaps 16;

QY 12 FSVLTFPAVSADKIPGDEITNIGPRDNSSPKHILNNHTAYSESHTLYDRMTFL 71  
 DB 293 YNNVLRNNVYSKN---DYGIAMLY-----SNNNVYNN--TASDNNRGILAVT-- 336  
 QY 72 CLSSHTNLNGACPTSENPSSSVSGEFTNITLOFTEKSLIKRELQIKGYKQLFKSVNCP 131  
 DB 337 --SSANFVSG-----KRAHSLNG--GIYLONCSNNNIISNEATLN-----TGGSTT 379  
 QY 132 SGLTINSA-HFNCKNAAGASLYLTPAGELKMLPFGIMDATLKLKRVKRRYSETGY 190  
 DB 380 NGIILGASNNVNNNIASENKRGITTSSSSGKRYSGN----- 418  
 QY 191 TINTITKLTIDKGNQIQLPQKSDARYDLNLPTG--GTYIGRNSVDMCFYDGYST-NS 247

Db	419	TLNSN--TGHC--IYLNAGTDNNLSNVASSNGVGIYL-VNSNNTALENNIATGNS	471
QY	248	SSLERFQDNNPKSDGKFLRKIN-DPTKEIAYTIS--LLAGKSLPTNGTSLNTADA	303
Db	472	KGIYVMTSNGNTISENEYNNNNVDVDSNGIMISLSNNKVSCKAYNNPYGISLNSSTN	531
QY	304	ASLETN	309
Db	532	NNISSN	537

Search completed: August 19, 2003, 09:48:13  
Job time : 105 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 19, 2003, 09:45:04 ; Search time 30 seconds  
(without alignments)  
509,141 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*

2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*

3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*

4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*

5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep.\*

6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17
2	988.5	52.4	363	2	US-08-483-101-16
3	927.5	49.2	364	2	US-08-483-101-5
4	116.5	6.2	1600	2	US-08-617-697-10
5	114	6.0	650	3	US-08-362-525-2
6	111.5	5.9	1005	4	US-09-206-942-41
7	111.5	5.9	1011	4	US-09-206-942-39
8	111.5	5.9	1529	2	US-08-728-470-10
9	111.5	5.7	524	4	US-09-242-913B-15
10	108	5.7	1095	4	US-09-206-942-45
11	107	5.7	1101	4	US-09-206-942-43
12	107	5.7	1101	4	US-09-206-942-43
13	106	5.6	671	2	US-08-737-716-13
14	101	5.4	915	4	US-09-206-942-35
15	101	5.4	1222	4	US-09-206-942-37
16	101	5.4	1228	4	US-09-206-942-34
17	101	5.4	1338	2	US-08-728-470-9
18	101	5.4	1338	3	US-08-719-641-9
19	101	5.4	1589	2	US-08-617-697-9
20	100.5	5.3	1004	4	US-09-206-942-57
21	100.5	5.3	1010	4	US-09-206-942-55
22	99.5	5.3	969	4	US-09-206-942-32
23	99.5	5.3	975	4	US-09-206-942-30
24	99	5.2	992	4	US-09-206-942-61
25	99	5.2	998	4	US-09-206-942-59
26	99	5.2	1848	3	US-08-296-791-6
27	99	5.2	1848	5	PCT-US95-10661A-6

28	97	5.1	977	4	US-09-206-942-53	Sequence 53, App1
29	97	5.1	983	4	US-09-206-942-51	Sequence 51, App1
30	94.5	5.0	1912	4	US-08-409-895-4	Sequence 4, App1
31	94.5	5.0	1912	3	US-08-685-467-4	Sequence 4, App1
32	94.5	5.0	2353	3	US-09-377-155-33	Sequence 33, App1
33	94.5	5.0	2353	3	US-08-913-942-4	Sequence 4, App1
34	94.5	5.0	2353	4	US-09-669-974-33	Sequence 33, App1
35	94.5	5.0	2354	4	US-09-268-347-47	Sequence 47, App1
36	94.5	5.0	2411	4	US-09-268-347-36	Sequence 36, App1
37	93.5	5.0	715	4	US-09-462-917A-134	Sequence 134, App
38	93.5	5.0	901	4	US-09-134-001C-5351	Sequence 5351, Ap
39	93	4.9	1167	2	US-08-485-568A-6	Sequence 6, App1
40	93	4.9	1167	2	US-08-590-554A-6	Sequence 6, App1
41	93	4.9	1167	2	US-09-184-223-6	Sequence 6, App1
42	92.5	4.9	990	4	US-09-252-991A-32469	Sequence 32469, A
43	92.5	4.9	1073	4	US-09-206-942-49	Sequence 49, App1
44	92.5	4.9	1079	4	US-09-206-942-47	Sequence 47, App1
45	91.5	4.9	869	2	US-08-483-101-15	Sequence 15, App1

## ALIGNMENTS

RESULT 1  
US-08-483-101-17  
Sequence 17, Application US/08483101  
Patent No. 5932715  
GENERAL INFORMATION:  
APPLICANT: Scott, June R.  
APPLICANT: Froehlich, Barbara  
TITLE OF INVENTION: CS2 Proteins and Coding Sequences  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,101  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 6-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-483-101-17  
Query Match 81.5%; Score 1536.5; DB 2; Length 360;  
Best Local Similarity 80.1%; Pred. No. 1e-149;  
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;  
QY 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 60  
Db 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 60

QY	61	SHLTLDRTPLCLSSHTNLNGACPSENNSSSVSEGTITLQFTKRSLLIRELOIKY	120
Dd	61	SHLTLDRTPLCLSSHTNLNGACPSENNSSSVSEGTITLQFTKRSLLIRELOIKY	120
QY	121	KOLFKSVNCPBGLTLLNSAHFNCKNNKNAASGASLYIYPAGELKNLPFGIMDATTLRYK	180
Dd	121	KOLFKSVNCPBGLTLLNSAHFNCKNNKNAASGASLYIYPAGELKNLPFGIMDATTLRYK	180
QY	181	RRSEYTYTYTINITYIKLTDKNIOIMLQFKSDARVDNLNRPFGGTYIGRNSVMCFY	240
Dd	181	RRDYTYTYTYTINITYIKLTDKNIOIMLQFKSDARVDNLNRPFGGTYIGRNSVMCFY	240
QY	241	DGYSTNSSSLEIRFODNNPKSDGKRYLKRINDITEIATLSLLAGKSLPTNGTSLNT	300
Dd	241	DGYSTNSSSLEIRFODNNPKSDGKRYLKRINDITEIATLSLLAGKSLPTNGTSLNT	300
QY	301	ADAALESNNNRITAVTMEPEISVPYLCWGRQLOLAKVNNPEAGGYMGINITFPPSSOR	360
Dd	301	ADAALESNNNRITAVTMEPEISVPYLCWGRQLOLAKVNNPEAGGYMGINITFPPSSOR	360
QY	361	L 361	
Dd	360	L 360	

```

1      RESULT 2
2      US-08-483-101-16
3      Sequence 16, Application US/08483101
4      Patent No. 5932715
5      GENERAL INFORMATION:
6      APPLICANT: Scott, June R.
7      APPLICANT: Froehlich, Barbara
8      APPLICANT: Caron, Judy
9      TITLE OF INVENTION: CS2 Proteins and Coding Sequences
10     NUMBER OF SEQUENCES: 17
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: Greenlee and Winner, P.C.
13     STREET: 5370 Manhattan Circle, Suite 201
14     CITY: Boulder
15     STATE: Colorado
16     COUNTRY: US
17     ZIP: 80303
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.30
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/483,101
25     FILING DATE: 07-JUN-1995
26     CLASSIFICATION: 424
27     ATTORNEY/AGENT INFORMATION:
28     NAME: Feibler, Donna M.
29     REGISTRATION NUMBER: 33878
30     REFERENCE/DOCKET NUMBER: 6-95
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (303) 499-8080
33     TELEFAX: (303) 499-8089
34     INFORMATION FOR SEQ ID NO: 16:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 363 amino acids
37     TYPE: amino acid
38     STRANDEDNESS:
39     TOPOLOGY: unknown
40     MOLECULE TYPE: protein
41     HYPOTHEICAL: NO
42     US-08-483-101-16

```

```

QY      3 KILFIITLFPSSVLFPFAVSADKIPGDE--SITNIF-GPR-DRNESSPKRIINHTTAV 58
D      2 KKIIFIL-----SIFSAVVSAGRIPEITTYGNLTTSFQAPLDRBSQSFYIIFINHVAGY 57
QY      59 SESHTLYDRMTFLCISSHNTLNGACPTSENPSSSVS-GETNITLQFTKRSKLRELOI 117
D      58 SLHSILDRIVFLCTSSSPNVGACPTI---GTSQVQGTITTLITQFTKRSKLRIINIL 114
QY      118 KGYKQILFESVNCPSG--LTLNAPFNCKNA-ASGASLYITPAGELKULPREGIYDAT 174
D      115 AGKKRKRIMENOSCDFSNLVLNLSKMSGCAHNAAGTLNLNIYPAGEINKLIPFGIWEAT 174
QY      175 LKLRVRRRYSET-----YGYTTINITIKTLDKGNIOIMLPQKSPARDNLNRPGGGTY 230
D      175 LIIRLS-RIGEVSSITHYGNITVINITVLDLKGNIOWMLPGHSPRVDNLNRPGLGNTKYS 233
QY      231 GRNSVDMCFYDGYSTNSSSLERFODNNPKSDGKEYELKINDDPKREIATYLLSLLAGSL 290
D      234 GNSLDMCFYDGYSTNDSDMVIKFPQDDNPNTNSEVENLKIGI-GEKLPYAVNSILMGKIF 292
QY      291 TPIPNGLSLNADAALEIWNRIITVYIMPEISVPLYLCPGRLQIDAKAYENPEAGQIYGN 350
D      293 YPNAGOSGFIINDSSVLEYETIMNRVTAYVAMBEVAVPLYLCPARLLNADVNPBEAGQYGN 352
QY      351 NNTFPPSSQTL 361
D      353 KIITFPPSSQTL 363

```

```

: RESULT 3
: US-08-483-101-5
: Sequence 5, Application US/08483101
: Patent No. 5932715
: GENERAL INFORMATION:
: APPLICANT: Scott, June R.
: APPLICANT: Enoehlich, Barbara
: APPLICANT: Caron, Judy
: TITLE OF INVENTION: CS2 Proteins and Coding Sequences
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee and Winner, P.C.
: STREET: 5370 Manhattan Circle, Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: US
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,101
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Feiber, Donna M.
: REGISTRATION NUMBER: 33878
: REFERENCE/DOCKET NUMBER: 6-95
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 499-8080
: TELEFAX: (303) 499-8089
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 364 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-483-101-5

```

Query Match	52.4%	Score 988.5	DB 2:	Length 363;
Best Local Similarity	55.0%	Pred. No. 2.9e-93;		
Matches 204; Conservative	53;	Mismatches 93;	Indels 21;	Gaps 11.

Query Match 49.2%; Score 927.5; DB 2; Length 364;  
Best Local Similarity 48.1%; Pred. No. 5.5e-87;  
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

```

OY 1 MNKLEIFLTFESSVLETFEAVSADKIPGDEITNIGPRDRN--ESSPKNNIINNITAY 58
DB 1 LKAVLIVLSMFLCSOYTGOSMHTNVEAGINKTESIGPLDRSAAASYPAHYLFHEHVAGY 60
OY 59 SESHTLYDMRTFLCLSSHNTLNGACPTSENPSSSVSGEINTLTOTTERSLIKRELQIK 118
DB 61 NNDHSLFDMRTFLCLSSHNTLNGACPTSENPSSSVSGEINTLTOTTERSLIKRELQIK 118
OY 119 GKOLLEFKSVNC---PSGLTNSAHFNCKNAASGASLYLYIPAGELKMLPPGQWDAATL 175
DB 119 GKRFLEYESDRCHYVDKNNLSHFKVCYSGFTREGDFTLYIPOGEIDLLTGCIWEATL 178
OY 176 KLRVRKRYSETGYTYINTIKLTDKNGIOIMLPKPSDARDVNLNRPFGGTYIGRNVY 235
DB 179 ELRVKRYHYNMGTIVNITVDLTDKNGIOIMLPKPSDARDVNLNRPFGGTYIGRNVY 238
OY 236 DMCFTDYGSYTNSSSLEIRFQDNNPKSGKRYLKRINDTKELAYTLSELLAGSLPTNG 295
DB 239 EMCLYDGYSTHSGSIEMRFQDSDQGTGNNEYNLIKTGEPLKLPYKLSLLGGREFYPNG 298
OY 296 TSLNTADASLETNNRITAVTMEPEISVYLCMPGRLQDAVENEPAQOYMGINIVT 355
DB 299 EAFITNDYSSLFINMNRKISVSLPOLISIVLCMPANLTEMSELNPEAGEYSGILNVTET 358
OY 356 PSSQTL 361
DB 359 PSSSSL 364

```

```

RESULT 4
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336

```

```

GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617, 697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424

```

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651

```

```

REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813

```

```

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-10
Query Match
Best Local Similarity 21.6%; Score 116.5; DB 2; Length 1600;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;
OY 48 HNILNNHTAYSESHLYRMFTFLCSSHNTLNGACPTSENPSS---SYSGEINTLT 102
DB 816 YNEYSKH--AINSSHNL-----TTLGNTVLGG-----ENSSSSITGININTNKNVTL 862
OY 103 OF-----TEKRLIKRELQIKGYKOLLEFKSVNCPGSLT-----NSAHFNCON--- 144
DB 863 QADTSNNGSLKRLITLGNISVEGNSLTGANANIVGSLTAEDSTFGEASDNINITG 922
OY 145 ---KNAASGASLYLYIPAGELKMLPPGIMDAATLKLVRKRYSETGYT-YTINITIKLND 200
DB 923 TETNNGTANINIKGVYKLDGINNK--GC-----LNTTNASGTQKTIINGNIT--NE 970
OY 201 KGNIOIMLPKPSDARDVNLNRPFGGTYIGRNVYDMPGFTDYSTNSSSLEIRQ--- 255
DB 971 KQDLNT--KNIRADAEIQGNSISQEGMLTSSDKVNT-----TNOITKAGVEGGR 1021
OY 256 -DNNPKSDGKFLYLRKINDTKELIAYTLSELLAG---KSLPTNGTSLNTADASLETNN 311
DB 1022 SDSESENNANLTIQ-----TKELKLAGDLNIGFKNKAEITAKNGSDTLTGMSGCGNADAK 1076
OY 312 RITAVTMEPEISVYLCMPG-RLQDAVENEPAQOYMGINIVT 353
DB 1077 K--VFEDKVKDSKISTDGHNTLNSEVTSNGSSNAGNDNST 1116

```

```

RESULT 5
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910

```

```

GENERAL INFORMATION:
APPLICANT: KLIIS, FRANCISCUS M.
APPLICANT: SCHREIDER, MAARTEN P.
APPLICANT: TOSCHKA, HOUZER Y.
TITLE OF INVENTION: VERRIPS, CORNELIS T.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5918

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,525
FILING DATE: 04-JAN-1995
CLASSIFICATION: 435

```

```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992

```

```

ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 92203899.7

```

```

REFERENCE/DOCKET NUMBER: 92203899.7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813

```

```

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid

```

```

?      REGISTRATION NUMBER: 16,773
?      REFERENCE/DOCKET NUMBER: 213289/77020(V)
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (202) 861-3000
?      TELEFAX: (202) 822-0944
?      TELEEX: 6714627 CUSH
?      INFORMATION FOR SEQ ID NO: 2:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 650 amino acids
?      TYPE: amino acid
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
US-08-362-525-2

```

Query Match	6.0%;	Score 114;	DB 3;	Length 650;
Best Local Similarity	20.9%;	Pred. No. 0.0085;		
Matches 98;	Conservative 52;	Mismatches 147;	Indels 172;	Gaps 23

44 SSP---KHNILNNHITA-----YSESHLYDRMTFLC---LSSHNTLN 80  
Dj 68 SMFHVRIKLNNSSCATATSLADCTEAFKCYVSCQAALYENTFETCTCAQNDLSSNYTID 127  
Qy 81 GACPTSENNSSSSVSGEITITLQTFEKSILKRELIQIGYQQLKYSNC-PSGLTJNSA 139  
128 GSTTFEFLNSDGGSSTEYEL-----ENAKFEKSGPMALIKLQNMOSDVNEDPAFTNVF 182  
Dj 140 HFNCNKNASGASLYL-YIPAGELKNLPFGIMPATLKLHVKKRRSEYTYATYINTIK 197  
Qy 183 HSGRSIGYSGPSFESHGMYCPRNGYF---LGS---TEKI-----DIVSSNNNDL- 225  
Dj 198 LTRKGIQI-----WLPQFKDARVDLNRPTGGGITYGRN---SYDMCTYDG--- 242  
Qy 226 --DQSSQVYSSNDENDWMFPOSYNDTNADV-----TQFGSNLMTLDEKLYDEML 275  
Dj 243 -----YSTNSSLE--IFRODNN--PKSDGKF 265  
Dj 276 WVAALOSLPANNTIDHALFQYTCLODITANTTYATQPSITRRETYVYQGRNLGINSKSS 335  
Qy 266 YLRKINDDTKEI--AYTLS-----LLDAGSLPPTNGTSLINADAA 304  
Oy 336 FISTTTTDLTSLNTSAVSTGSISETVEGNRTTSEVISHVVTSTKSLPFAATSLTIQOTS 395  
Dj 305 SLETMNN-----RIAYVMPELSYVLCWPGRL 332  
396 IYSTDSNITVGTDIHTTSEVISOVETISIRASTVVAAPUSFTGWTGM 444

RESULT 6 -942-41  
 US-09-206-942-41  
 : Sequence 41. Application US/09206942  
 : Patent No. 6432869  
 : GENERAL INFORMATION:  
 : APPLICANT: Loosmore, Sheena M.  
 : APPLICANT: Yang, Yan-Ping  
 : APPLICANT: Klein, Michel H.  
 : TITLE OF INVENTION: Protective Recombinant Haemophilus  
 : TITLE OF INVENTION: Molecular Weight Proteins  
 : FILE REFERENCE: 1038-861 MIS:jb  
 : CURRENT APPLICATION NUMBER: US/09/206,942  
 : CURRENT FILING DATE: 1998-12-08  
 : EARLIER APPLICATION NUMBER: 09/167,568  
 : EARLIER FILING DATE: 1998-10-07  
 : NUMBER OF SEQ ID NOS: 95  
 : SOFTWARE: Patentin Ver. 2.1  
 : SEQ ID NO 41  
 : LENGTH: 1005  
 : TYPE: PRF  
 : ORGANISM: Haemophilus influenzae  
 : US-09-206-942-41

Query Match	5.9%;	Score 111.5;	DB 4;	Length 1005;
Best Local Similarity	20.3%;	Pred. No. 0.031;		
Matches 68;	Conservative 65;	Mismatches 147;	Indels 55;	Gaps 14;

```

OY      60  ESTLTDRTMPTCLTSSHTNLNACPTSEPPSSVSGCTNTLTQFTEKRS--LIRREJQIK 118
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      363  EKRAITSTHNLTLTGNVTTLGSESSSNTKGINININSKANVTLOAHAGTSHLDKE---- 418
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      119  GYKOLFFKSVNCPSCGTLT--NSAHFNCKMNAASGASLYLTPAGEL--KNLPFGGIV--DA 173
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      419  --RTLLGNVSVGNNLNTIGSNAHLDGNLSIAESKF-----QCKTNNNLNTCTFTNNG 471
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      174  TLKLRVRRRYSEYGTYYTINTPTIKLTD-----KGNV-----OIWLPQFSKARDV 218
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      472  TADINIKQGVKALQGGITNNGNLNTTNASVWQKTIINGNTTNKKGDNLKIDIKANEIQ 531
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      219  L--NLRPFGGTYIGANSYDMCFYDGYSTSSSLEIRPQDNNPKSD--GKFLYRKINDTK 275
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      532  IGENISQKEGNLTLISSDKINI-----TKRIEIKADTLQDQSSDGVSANMLTITKT 582
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      276  ELIVYTLSSLLAG---KSLLPTMGTSLNLTADAASLETNNRTATYATMEISVPLCWGR 332
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      583  ELTLTNNLINSGFNKAETIAKNSDLIIIGKASSDNSNAKQITFEDKVRKDSKIS--AGNHNV 640
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY      333  OLDKAVENPEA-----GOYMGNIWVTFPPSQTL 361
           | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db      641  TLNSKVEITSNDGSTGNGSDNNNIGLTLISAKDYV 675

```

RESULT 7  
 US-09-206-942-39  
 Sequence 39, Application US/09206942  
 Patent No. 643269  
 GENERAL INFORMATION:  
 APPLICANT: Loosmore, Sheena M.  
 APPLICANT: Yang, Yan-Ping  
 APPLICANT: Klein, Michael H.  
 TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
 TITLE OF INVENTION: Molecular Weight Proteins  
 FILE REFERENCE: 1038-861 MIS:jb  
 CURRENT APPLICATION NUMBER: US/09/206,942  
 CURRENT FILING DATE: 1998-12-08  
 EARLIER APPLICATION NUMBER: 09/167,568  
 EARLIER FILING DATE: 1998-10-07  
 NUMBER OF SEQ ID NOS: 95  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 39  
 LENGTH: 1011  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-09-206-942-39

	Query Match	Best Local Similarity	Score	DB 4	Length	Indels	Gaps
			5.9%		111.5		
			20.3%		Pred. No.	0.031	
	Matches	68;	Conservative	65;	Mismatches	147;	Indels 55; Gaps 14
QY	60	ESHTLYDRMTFLCSHNTLNGACPTSENPSSSVSGETNITLQFTKRS-LIKREIQIK	118				
Db	369	EKAHISTNLTILGCVNTILGESSNSIKINININSKANVTLOAHAGTSHLDKE----	424				
QY	119	GYNOLFFKVCNPGSLT--NSAHCNCNKMAASGLIYIPAGEL-KNPFGIW--DA	173				
Db	425	--RTLLGVAVSGGNINIIGSNAHLIDGLIAESAKF-----QGKTNNNLNITGTFPNNG	477				
QY	174	TTLRYRRYSSEYGYTNITIKLTD-----KGNT-----OIWPFQSDARD	218				
Db	478	TADINKQGVVLQGDITFNGMLNTTTNAAVNOKITIIINGNITTKGGDLTIKDIAVAEIQ	537				
QY	219	L-NLRPFGGGYIIGNSVDMCFYCYGSTSNSSLEIRQDNNPKSD-CRKYLRKINDPK	275				
Db	538	IGGNISQKEGNLTJSSDKINTI-----FKRIEKADDVDGSSDGVSANANLTITKY	588				



```
QY      276 ELIAYTSLILAG---KSLPTMGSTGLNTLADAASLETENNRRITTAATMEISVPLWCGRL 332
Db      589 ELFTLTNLNIGSGFNKAETAKDNPDLIIIGKASSDNMSNAKQTTFPKVADSKIS--AGHHNV 646

QY      333 QLDAAVEPEA-----GQYMGINVTETPSSQTL 361
Db      647 TLNSKVEITSNDOSTGNGSDDDNNIGITLSAKDYTY 661
```

RESULT 8  
US-08-728-470-10

Sequence 10, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteinms  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington, Va 22202-4302

STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CHIPPED: Application forms

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/728,470
3 FILING DATE:
4 CLASSIFICATION: 424
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US 08/302,832
7 FILING DATE: 16-MAR-1993
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: US PCT/US93/02166
10 FILING DATE: 16-MAR-1993
11 PRIOR APPLICATION DATA:

```

1 APPLICATION NUMBER: GB 9205704.1  
2  
3 FILING DATE: 16-MAR-1992  
4  
5 ATTORNEY/AGENT INFORMATION:  
6 NAME: Beiklesses, Jerry M  
7  
8 REGISTRATION NUMBER: 22,651  
9  
10 REFERENCE/DOCKET NUMBER: 1038-6533  
11  
12 TELECOMMUNICATION INFORMATION:  
13 TELEPHONE: (703) 415-0810  
14  
15 TELEFAX: (703) 415-0813  
16  
17 INFORMATION FOR SEQ ID NO: 10:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-728-470-10

```

Query Match	5.9%	Score	11.5;	DB	2;	Length	1529;
Best Local Similarity	20.7%	Pred. No..	0.061;				
Matches	72;	Conservative	62;	Mismatches	127;	Indels	87;
						Gaps	18

```
OY      48 H N I L N H T A Y S E S H T L D R M F L C L S N I L N G A C P T S N P S S -----SVSGEINTTL 102
```

DB	744	YNEXSKH--AINSSHNL-----TILGNVTLGG-----ENSSSITGININITKANVIL	790
OY	103	QF-----TEKRSLIKRELOINGYKOLLFYSVNCPSGITLTNSAHFNCKNMAASGASLY	154

D**b**  
791 QADTSNSMTGLKRRLLTIGNISVYGNNLSITGANANIVGNLSI--AEDSFFKEAS----- 843

155 LYPAGELKNLPPGCIW--DATIKLRVKRRISEYGYT---INITIKLT----- 199

Db 844 -----DNLNTGTEFFNNGTANINIKQVVKLQGGIDINNKGGINITTNAGSGFTQTING 8955

Qy 200 ----DKGNIOIOLPQFESDARVLD--NLRPTGGGTYIGRNVDCMCFDGYSTNSSSEIR 2533

896 NITNEGGDLNT--KNIKADAEIOIGNSICEKGNITLISDKVNI-----TNDITTKAG 9466

Db 254 FO----DNNPKSDGKPYLKRINDTKEETAYVLSTLLAG--KSLTPNGTSLNIADASTL 3066

QY 307 ETNNNRITAVT<sup>1</sup>MPETSYEVLCPG-R<sup>2</sup>LQ<sup>3</sup>LDAKVENPEAGQYMG<sup>4</sup>NI<sup>5</sup>NT 353

```

Db      1002 NADAKK---VFEDKVKDSKISFDGHNNVLTINSEVKTSSNGSSNMGNDNST 1

RESULT 9
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

1 PAKISTAN AFFIDAVIT DATA:  
 2 APPLICATION NUMBER: US 08/302,832  
 3 FILING DATE: 16-SEP-1994  
 4 PRIOR APPLICATION DATA:  
 5 APPLICATION NUMBER: US PCT/US93/02166  
 6 FILING DATE: 16-MAR-1993  
 7 PRIOR APPLICATION DATA:  
 8 APPLICATION NUMBER: GB 9205704.1  
 9 FILING DATE: 16-MAR-1992  
 10 ATTORNEY/AGENT INFORMATION:  
 11 NAME: Barstrower, Terrv W

REGISTRATION NUMBER: 22,665  
REFERENCE/DOCKET NUMBER: 1  
TELECOMMUNICATION INFORMATION  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10

```

; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
;

```

```

      !          TOPOLOGY:  linear
      US-08-719-641-10

```

Query Match	5.9%;	Score 11.5;	DB 3,	Length 1529;
Best Local Similarity	20.7%;	Pred. No. 0.061;		
Matches 72;	Conservative 62;	Mismatches 127;	Indels 87;	Gaps 16

48 HNIILNNHTAYSSEHTLYDRMTFLCLSSHNHTLNGACPFSNPPSS-----SVSGFTNITL 102

Db 744 YNEYSKH--AINSSHL-----TILGGAVTLGC-----ENSSSITGNTNITNKANVTL 790





Db 489 GEVKSASGNVNTASGNTLNISNITGQVTVTAN-----SGAITTEGST----- 533  
QY 332 LQIDAKVENPEAGQYMGNNIN-----VFTEPSSQTL 361  
Db 534 --INATTDGANITTTGTGNGKVESSSGSVTLIATGQTL 570

## RESULT 15

US-09-206-942-37  
; Sequence 37, Application us/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
FILE REFERENCE: 1038-861 MIS:jb  
CURRENT APPLICATION NUMBER: US/09/206,942  
CURRENT FILING DATE: 1998-12-08  
EARLIER APPLICATION NUMBER: 09/167,568  
EARLIER FILING DATE: 1998-10-07  
NUMBER OF SEQ ID NOS: 95  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 37  
LENGTH: 1222  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-206-942-37

Query Match 5.4%; Score 101; DB 4; Length 1222;  
Best Local Similarity 21.1%; Pred. No. 0.51;  
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIPGDESDTNTFFGR-----DRNESSPKH-----NILN-----NHITAYS 59  
Db 559 ISSDKNVNTERTITKAGVNGSDNSNEATSAFLTITKELKLTNDLNSGENKAEITAKD 618  
QY 60 ESH-TLYD-----RMTF-----LCISSHN-TLNGACPTSENPSSSSVSGETNI 100  
Db 619 NSNLIGNSDAGNTDAKVFNSNKKDSIASDHNVTLNSKVEISGDTSTEDGNNNT 678  
QY 101 TLQFTKRSLLKRELQIKYQOL--LFKSVNCPSEGLTNSAHFNCKNNAASGASLYLYI 157  
Db 679 GLTITAKNVTVNN--NITSKTVNITASENVTTKAGTTIN-----ATTGSVEVTA 726  
158 PAGELKNLPFGIMDATLKLKVKRYSFTYTTITITIKLTKDNGNIQIWLPOFKSDARV 217  
Db 727 KTGDIK--GGI-----ESNSGNVNI----- 744  
QY 218 DLNLEPTGCGTYI-----GRNSVDMCFYDGYSTNSSLEIRFQDNPKSDGKFLRKIN 271  
Db 745 -----TASGDTLANSNITGQV--VVAASGAVTTTKGSTINATGNNANITTK--TGEIN 795  
QY 272 DDTKRIATYLLSLLAGKSLITPLNGSLNIADASLETMNNRTTAVTMEISVPLCWPGR 331  
Db 796 GEVKSASGNVNTASGNTLNISNITGQVTVTAN-----SGAITTEGST----- 840  
QY 332 LQIDAKVENPEAGQYMGNNIN-----VFTEPSSQTL 361  
Db 841 --INATTDGANITTTGTGNGKVESSSGSVTLIATGQTL 877

Search completed: August 19, 2003, 09:49:35  
Job time : 32 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:48:19 : Search time 57 Seconds  
(without alignments)  
829.687 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886  
Sequence: 1 MNKILFFLFFSSVLTFFFA.....EAGQYMGNINVTFFSSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues

Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications-AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	100.0	361	US-09-839-894-10	Sequence 10, Appl
2	1862.5	98.8	361	US-09-839-894-28	Sequence 28, Appl
3	1542	81.8	359	US-09-839-894-29	Sequence 29, Appl
4	937	49.7	362	US-09-839-894-31	Sequence 31, Appl
5	916.5	48.6	364	US-09-839-894-30	Sequence 30, Appl
6	247.5	13.1	353	US-09-839-894-32	Sequence 32, Appl
7	119.5	6.3	770	US-09-303-232-2	Sequence 2, Appl
8	116.5	6.2	1600	US-10-092-880-10	Sequence 10, Appl
9	111.5	5.9	1005	US-10-193-764-41	Sequence 41, Appl
10	111.5	5.9	1011	US-10-193-764-39	Sequence 39, Appl
11	109	5.8	2383	US-09-912-020-302	Sequence 302, Appl
12	108	5.7	524	US-10-347-278-15	Sequence 15, Appl
13	101	5.4	915	US-10-193-764-35	Sequence 35, Appl
14	101	5.4	1222	US-10-193-764-37	Sequence 37, Appl
15	101	5.4	1228	US-10-193-764-34	Sequence 34, Appl

16	101	5.4	1599	US-10-092-880-9	Sequence 9, Appl
17	100.5	5.3	1004	US-10-193-764-53	Sequence 53, Appl
18	100.5	5.3	1010	US-10-193-764-51	Sequence 51, Appl
19	99.5	5.3	969	US-10-193-764-32	Sequence 32, Appl
20	99.5	5.3	975	US-10-193-764-30	Sequence 30, Appl
21	99	5.2	839	US-10-023-437-23	Sequence 23, Appl
22	99	5.2	992	US-10-193-764-57	Sequence 57, Appl
23	99	5.2	998	US-10-193-764-55	Sequence 55, Appl
24	99	5.2	1848	US-09-839-996-6	Sequence 6, Appl
25	99	5.2	1848	US-10-080-505-6	Sequence 6, Appl
26	97	5.1	977	US-10-193-764-49	Sequence 49, Appl
27	97	5.1	983	US-10-193-764-47	Sequence 47, Appl
28	95.5	5.1	825	US-10-101-464A-922	Sequence 922, Appl
29	95	5.0	867	US-09-839-894-6	Sequence 6, Appl
30	94.5	5.0	2353	US-09-797-862-33	Sequence 33, Appl
31	92.5	4.9	1073	US-10-193-764-45	Sequence 45, Appl
32	92.5	4.9	1079	US-10-193-764-43	Sequence 43, Appl
33	91.5	4.9	1143	US-09-924-154-14	Sequence 14, Appl
34	91.5	4.9	1236	US-09-769-787-109	Sequence 109, Appl
35	91	4.8	1395	US-10-080-505-7	Sequence 7, Appl
36	90.5	4.8	564	US-09-944-160-12	Sequence 12, Appl
37	90.5	4.8	1019	US-09-776-191-64	Sequence 64, Appl
38	90.5	4.8	1019	US-10-157-031-267	Sequence 267, Appl
39	89.5	4.7	811	US-10-011-588-29	Sequence 29, Appl
40	89	4.7	827	US-10-214-524-36	Sequence 36, Appl
41	89	4.7	465	US-09-957-485-6	Sequence 6, Appl
42	89	4.7	465	US-09-987-021-6	Sequence 6, Appl
43	88	4.7	921	US-09-117-447-6	Sequence 6, Appl
44	88	4.7	1220	US-10-193-764-28	Sequence 28, Appl
45	88	4.7	1226	US-10-193-764-26	Sequence 26, Appl

#### ALIGNMENTS

RESULT 1  
US-09-839-894-10  
Sequence 10, Application us/09839894  
Patent No. US20020176868A1  
GENERAL INFORMATION:  
APPLICANT: Alboum, Zeev  
APPLICANT: Barry, Eileen M.  
APPLICANT: Levine, Myron M.  
APPLICANT: University of Maryland  
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE  
FILE REFERENCE: USFMD.006A  
CURRENT APPLICATION NUMBER: US/09/839,894  
CURRENT FILING DATE: 2001-04-20  
PRIOR APPLICATION NUMBER: 60/198,626  
PRIOR FILING DATE: 2000-04-20  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 361  
TYPE: PRT  
ORGANISM: E. coli  
US-09-839-894-10

Query Match 100.0%; Score 1886; DB 10; Length 361;  
Best Local Similarity 100.0%; Pred. No. 1.5e+181;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKILFFLFFSSVLTFFFAADKIPGDESTINIGPRDRNESSPKHNLNNHTAYSE 60  
DB 1 MNKILFFLFFSSVLTFFFAVADKIPGDESTINIGPRDRNESSPKHNLNNHTAYSE 60  
QY SHTLYRMFLCLSSNNTNGACPTSENPSSSVSEGTITLOFTFKRLIKRELOIKY 120  
DB SHTLYRMFLCLSSNNTNGACPTSENPSSSVSEGTITLOFTFKRLIKRELOIKY 120  
QY KOLLFRVNCPSGLTLNSAHFNCNKAASGASLIYIPAGELKNLPFGGITADATKLRYK 180  
DB KOLLFRVNCPSGLTLNSAHFNCNKAASGASLIYIPAGELKNLPFGGITADATKLRYK 180

Db 121 KOLLEKSVNCPGSLTNSAHFNCKNNAASGASLYIIPAGELKNIPEFGIMDATIKLRVK 180  
QY 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 240  
Db 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 240  
QY 241 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 300  
Db 241 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 300  
QY 301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 360  
Db 301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 360  
QY 361 L 361  
Db 361 L 361

## Query 2

US-09-839-894-28  
; Sequence 28, Application US/09839894  
; Patent No. US20020176868A1  
; GENERAL INFORMATION:  
; APPLICANT: Alboum, Zeev  
; APPLICANT: Barry, Eileen M.  
; APPLICANT: Levine, Myron M.  
; APPLICANT: University of Maryland  
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE  
; FILE OF INVENTION: CSA OPERON  
; FILE REFERENCE: USFMD 006A  
; CURRENT APPLICATION NUMBER: US/09/839,894  
; CURRENT FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/198,626  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 361  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: EREC Protein Homology Sequence  
US-09-839-894-28

Query Match 98.8%; Score 1862.5; DB 10; Length 361;  
Best Local Similarity 99.4%; Pred. No. 3.6e-179;

Matches 359; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESTITNIFGRDRNESSPKHNIINHTAYSE 60  
Db 2 MNKILFIFTLFFSSVLTFFAVSADKIPGDESTITNIFGRDRNESSPKHNIINHTAYSE 61  
QY 61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITLQTFEKRSLIKRELOIKGY 120  
Db 62 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITLQTFEKRSLIKRELOIKGY 121  
QY 121 KOLLEKSVNCPGSLTNSAHFNCKNNAASGASLYIIPAGELKNIPEFGIMDATIKLRVK 180  
Db 121 KOLLEKSVNCPGSLTNSAHFNCKNNAASGASLYIIPAGELKNIPEFGIMDATIKLRVK 181  
QY 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 240  
Db 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 241  
QY 241 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 300  
Db 242 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 300  
QY 301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 360  
Db 301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 360

QY 361 L 361  
Db 361 L 361

## RESULT 3

US-09-839-894-29  
; Sequence 29, Application US/09839894  
; Patent No. US20020176868A1  
; GENERAL INFORMATION:  
; APPLICANT: Alboum, Zeev  
; APPLICANT: Barry, Eileen M.  
; APPLICANT: Levine, Myron M.  
; APPLICANT: University of Maryland  
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE  
; FILE OF INVENTION: CSA OPERON  
; FILE REFERENCE: USFMD 006A  
; CURRENT APPLICATION NUMBER: US/09/839,894  
; CURRENT FILING DATE: 2001-04-20  
; PRIOR APPLICATION NUMBER: 60/198,626  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: EREC Protein Homology Sequence  
US-09-839-894-29

Query Match 81.8%; Score 1542; DB 10; Length 359;  
Best Local Similarity 80.9%; Pred. No. 7.3e-147;  
Matches 292; Conservative 32; Mismatches 35; Indels 2; Gaps 2;

QY 1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESTITNIFGRDRNESSPKHNIINHTAYSE 60  
Db 1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESTITNIFGRDRNESSPKHNIINHTAYSE 60  
QY 61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITLQTFEKRSLIKRELOIKGY 120  
Db 61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITLQTFEKRSLIKRELOIKGY 120  
QY 121 KOLLEKSVNCPGSLTNSAHFNCKNNAASGASLYIIPAGELKNIPEFGIMDATIKLRVK 180  
Db 121 KOLLEKSVNCPGSLTNSAHFNCKNNAASGASLYIIPAGELKNIPEFGIMDATIKLRVK 180  
QY 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 240  
Db 181 RRYSETYGTNTITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGRNSVDMCFY 239  
QY 241 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 300  
Db 240 DGYSTNSSLEIRFODNNPKSDGKFKYLRKINDOTEIAYTSLLAGKSLPTNGTSLNI 299  
QY 301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 360  
Db 300 -NTASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIWVTFPPSSOT 358  
QY 361 L 361  
Db 359 L 359

## RESULT 4

US-09-839-894-31  
; Sequence 31, Application US/09839894  
; Patent No. US20020176868A1  
; GENERAL INFORMATION:  
; APPLICANT: Alboum, Zeev  
; APPLICANT: Barry, Eileen M.  
; APPLICANT: Levine, Myron M.  
; APPLICANT: University of Maryland

Query Match	49.7%;	Score 937;	DB 10;	Length 362;
Best Local Similarity	53.1%;	Pred. No. 7.5e-86;		
Matches 197; Conservative	54;	Mismatches 98;	Indels 22;	Gaps 12

SULT 5  
8-09-839-894-30  
Sequence 30, Application US/09839894  
Patent NO. US20020176868A1  
GENERAL INFORMATION:

```

APPLICANT: University of Maryland
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
TITLE OF INVENTION: CSA OPERON
FILE REFERENCE: UOFMD.006A
CURRENT APPLICATION NUMBER: US/09/839,894
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/198,626
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ. ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO 30
LENGTH: 364
TYPE: PRN
ORGANISM: Artificial Sequence
FEATURE:

```

Query Match	48.68;	Score	916.5;	DB	10;	Length	364;
Best Local Similarity	48.18;	Pred. No.	8.8e-84;				
Matches	176;	Conservative	63;	Mismatches	120;	Indels	7;
						Gaps	3;

RESULT 6  
US-09-839-894-32  
; Sequence 32, Application US/09839894

```

? GENERAL INFORMATION:
? APPLICANT: Alboum, Zeev
? APPLICANT: Barry, Eileen M.
? APPLICANT: Levine, Myron M.
? APPLICANT: University of Maryland
? TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
? TITLE OF INVENTION: CSA OPERON
? FILE REFERENCE: UOFMD, 006A
? CURRENT APPLICATION NUMBER: US/09/839,894
? CURRENT FILING DATE: 2001-04-20
? PRIOR APPLICATION NUMBER: 60/198,626
? PRIOR FILING DATE: 2000-04-20
? NUMBER OF SEQ ID NOS: 40
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 32
? LENGTH: 353
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: ETEC Protein Homology Sequence
? US-09-839-894-32

```

Query Match	13.1%	Score 247.5	DB 10	Length 353
Best Local Similarity	28.9%	Pred. No. 2,4e-16		
Matches	89	Conservative 42	Mismatches 126	Indels 51
				Gaps 13
OY	77	NTLNG-ACPTSENSSSS-----VSGETNITLQPTKEKSLIKREQLQIKGYKQLLEK	126	
		: : :   :   :   :   :		
Db	72	NVLGGWCRSRNNENEGCEETHLVWVYFGAYSILRREQISHAETL-----	120	
OY	127	SVNCPGGLTNSAHENC-----NKNNA---GASILTYIPAGELKMLPGGIMDÄTLKL-R	178	
		: : :   :   :   :   :		
Db	121	-----ILLGSVRDAGTGVINNNAAQCMQGSLLRIPSEELAKIPISGWTAKATLVLDY	173	





219 L--NRKPIUGGIIIGRNSVDMCFYDGISSINSSLEIKFQNNPKSD--GKFXLRKINDTK 2/3

DB -1080 GNGVDETLTATVKDP-SNHPVAGITVNF 1108

## RESULT 12

US-10-347-278-15

; Sequence 15, Application US/10347278

; Publication No. US20030147856A1

; GENERAL INFORMATION:

; APPLICANT: FONTAINE, THIERRY

; APPLICANT: HARTLAND, ROBERT

; APPLICANT: MOYNA, ISABELLE

; APPLICANT: LATGE, JEAN-PAUL

; TITLE OF INVENTION: METHOD FOR SORTING ANTIFUNGAL MOLECULES ACTING ON THE

; FILE REFERENCE: 05986-0007

; CURRENT APPLICATION NUMBER: US/10/347,278

; CURRENT FILING DATE: 2003-01-21

; PRIOR APPLICATION NUMBER: PCT/FR97/01540

; PRIOR FILING DATE: 1997-08-29

; PRIOR APPLICATION NUMBER: 60/024,910

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 15

; LENGTH: 524

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-10-347-278-15

Query Match 5.7%; Score 108; DB 12; Length 524;

Best Local Similarity 24.7%; Pred. No. 0.051; Mismatches 96; Indels 72; Gaps 18;

Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

QY 119 GYKOL--LFKSYNCPGSLTNSAHFNCKNNA---SGASLYLYIPAGELKMLPFGGIWDA 173

DB 264 GYDKLNSFEDVAVIP---LIFSEYGCNKNMPTREDEVSEGLY---GGLKNVSSGL--- 313

QY 174 TLKLVKRRYSEYTYGTITINITIKLTDKGNIOIMLPQKSD--ARYDLNR---PTGGGT 228

DB 314 -----VYETTEANNKGL---VKLDDSGSL-----TYKDDVNLIESQLKNVSLPTYES 359

QY 229 YIGRNSVDMC-----FYDGYSTNSSSLE-----IREQDNNPKSDGK----- 264

DB 360 EISDSIYKCDNSAITNIYSGGTNNFTLPSPAEIANNIEYGVNGTWT--GKILDDYAVP 418

QY 265 ----FYLRKINDTEIAYTSLSLAGSLPTPTNGTSLNIADAASLETNNMR--ITAVTMP 319

DB 419 TTFNTITNNKDDT--ISATISYDRA--NSLNELDVATATVAASASTOSSSRSLTSSTSP 475

QY 320 EISVPLVLCMPGRLQDAKVENPEAGQYMGNI--NVTFTPS 357

DB 476 SSSSTGSSSSTG-----SSSASSSSKKGVGNIVNVSFSQS 510

; Sequence 35, Application US/10193764

; Publication No. US20030133943A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH

; FILE REFERENCE: 1038-1239MIS

; CURRENT APPLICATION NUMBER: US/10/193,764

; CURRENT FILING DATE: 2002-07-12

; PRIOR APPLICATION NUMBER: 09/167,568

; PRIOR FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 35

; LENGTH: 915

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-193-764-35

Query Match 5.4%; Score 101; DB 12; Length 1222;

Best Local Similarity 21.1%; Pred. No. 0.61; Mismatches 129; Indels 138; Gaps 19;

Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIFGDESTITNIFGR---DRNESSPKH-----NILN-----NHITAYS 59

DB 252 ISSDKVNITERITIKAGVNGDSNENATSAVLITKTELKLTNDLINSGFNKAETAKD 311

QY 60 ESH-FLYP-----RMTF-----LCLSHN--TLNGACPTSENPSSSVSGEINI 100

DB 619 NSNLTIGNSDAGNTDAKVTFFSNVSKDISASDNVTLNSKVFESGJTDSTEDCGNNT 678

QY 101 TLOFTEKRSLLIKRELQIGYKOL--LFKSYNCPGSLTNSAHFNCKNNAASGASLYLYI 157

DB 679 GLITIAKAVYVNN--NITSHKTVNTTASENVYTKAGTIN-----ATGSEVEYTA 726

## US-10-193-764-35

Query Match 5.4%; Score 101; DB 12; Length 915;

Best Local Similarity 21.1%; Pred. No. 0.61; Mismatches 129; Indels 138; Gaps 19;

Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIFGDESTITNIFGR---DRNESSPKH-----NILN-----NHITAYS 59

DB 252 ISSDKVNITERITIKAGVNGDSNENATSAVLITKTELKLTNDLINSGFNKAETAKD 311

QY 60 ESH-FLYP-----RMTF-----LCLSHN--TLNGACPTSENPSSSVSGEINI 100

DB 312 NSNLTIGNSDAGNTDAKVTFFSNVSKDISASDNVTLNSKVFESGJTDSTEDCGNNT 371

QY 101 TLOFTEKRSLLIKRELQIGYKOL--LFKSYNCPGSLTNSAHFNCKNNAASGASLYLYI 157

DB 372 GLITIAKAVYVNN--NITSHKTVNTTASENVYTKAGTIN-----ATGSEVEYTA 419

QY 158 PAGELKMLPFGGIWDAITKLRYKRYSETTYGTITINITIKLTDKGNIOIMLPQKSDARV 217

DB 420 KTGDIK-----GGI-----ESNSGNVNI----- 437

QY 218 DLNLRPTGGGYI-----GRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSDGKFLRKIN 271

DB 438 -----TASGDTLNSNITGON--VTVAASGAVITTKGSLTINATGANNITTK--TGBIN 488

QY 272 DTFKEIAYTSLSLAGSLPTPTNGTSLNIADAASLETNNMRITAVTMPDISVPLVCMGR 331

DB 489 GEVKSASGNVNTASGNLTNSNITGQVNTYAN-----SGATITTEGST----- 533

QY 332 LQDAKVENPEAGQYMGNI-----VFTFPSSQTL 361

DB 534 --INATGTADANTTGTGNGINKEVSSSGSVTLIATGQL 570

; Sequence 37, Application US/10193764

; Publication No. US20030133943A1

; GENERAL INFORMATION:

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH

; FILE REFERENCE: 1038-1239MIS

; CURRENT APPLICATION NUMBER: US/10/193,764

; CURRENT FILING DATE: 2002-07-12

; PRIOR APPLICATION NUMBER: 09/167,568

; PRIOR FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 37

; LENGTH: 1222

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-10-193-764-37

Query Match 5.4%; Score 101; DB 12; Length 1222;

Best Local Similarity 21.1%; Pred. No. 0.96; Mismatches 129; Indels 138; Gaps 19;

Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIFGDESTITNIFGR---DRNESSPKH-----NILN-----NHITAYS 59

DB 559 ISSDKVNITERITIKAGVNGDSNENATSAVLITKTELKLTNDLINSGFNKAETAKD 618

QY 60 ESH-FLYP-----RMTF-----LCLSHN--TLNGACPTSENPSSSVSGEINI 100

DB 619 NSNLTIGNSDAGNTDAKVTFFSNVSKDISASDNVTLNSKVFESGJTDSTEDCGNNT 678

QY 101 TLOFTEKRSLLIKRELQIGYKOL--LFKSYNCPGSLTNSAHFNCKNNAASGASLYLYI 157

DB 679 GLITIAKAVYVNN--NITSHKTVNTTASENVYTKAGTIN-----ATGSEVEYTA 726



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:36:34 ; Search time 44 Seconds  
(without alignments)  
1302.279 Million cell updates/sec

Title: US-09-839-894-10  
Perfect score: 1886  
Sequence: 1 MKRKILFIETFFSVLTFPA.....EAGQYMGINIVTFPSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	1886	100.0	361	23	AAM50343
2	927.5	49.2	364	20	AAV22326
3	275.5	14.6	359	22	AAV45919
4	119.5	6.3	770	21	AAV50814
5	117.5	6.2	1386	24	ABR82573
6	117.5	6.2	1449	24	ABR82570
7	114	6.0	650	15	AAK47575
8	111.5	5.9	1005	21	AAAB01833
9	111.5	5.9	1011	21	AAAB01832

10	110.5	5.9	1529	14	AAK41732
11	110.5	5.9	1601	18	AAW30292
12	109	5.8	2383	21	AAK15945
13	107	5.7	1095	21	AAAB01835
14	107	5.7	1101	21	AAAB01834
15	106	5.6	671	17	AAAB01830
16	106	5.6	1222	21	AAAB01830
17	106	5.6	1228	21	AAAB01828
18	105	5.6	598	23	ABK47313
19	104.5	5.5	1440	23	ABK54801
20	101	5.4	1358	23	AAK41731
21	101	5.4	1598	18	AAW30291
22	100.5	5.3	1004	21	AAAB01841
23	100.5	5.3	1010	21	AAAB01840
24	99.5	5.3	969	21	AAAB01827
25	99.5	5.3	975	21	AAAB01826
26	99	5.2	839	23	ABP56002
27	99	5.2	839	24	ABU66267
28	99	5.2	839	24	ABU66267
29	99	5.2	992	21	AAAB01843
30	99	5.2	998	21	AAAB01842
31	97	5.1	418	21	AAK30907
32	97	5.1	454	21	AAK30906
33	97	5.1	957	21	AAAB01839
34	97	5.1	963	21	AAAB01838
35	96	5.1	430	21	AAK37667
36	96	5.1	513	21	AAK39277
37	96	5.1	631	21	AAK39276
38	96	5.1	683	21	AAK39275
39	95.5	5.1	825	21	AAK25547
40	95	5.0	867	23	AAM50341
41	95	5.0	2334	23	AAK31849
42	94.5	5.0	567	20	AAV37170
43	94.5	5.0	665	22	ABK68284
44	94.5	5.0	2353	17	AAK93933
45	94.5	5.0	2411	21	AAK23860

#### ALIGNMENTS

RESULT 1	
AAM50343	
ID	AAM50343 standard; Protein; 361 AA.
AC	AAM50343;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	ETEC CS4 pilus CsaE tlp associated protein.
XX	
KW	CS4 pilus; enterotoxigenic; ETEC; csa operon; CsaA; fimbrial;
XX	
OS	Escherichia coli.
XX	
FH	
FT	Key
FT	Peptide
FT	Protein
FT	Location/Qualifiers
XX	1..23
XX	/label= Signal_peptide
XX	24..361
XX	/label= Mature_protein
PN	WO200181582-A2.
XX	
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12914.
XX	
PR	20-APR-2000; 2000US-198686P.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Altbaum Z, Levine MM, Barry EM;

High molecular weight  
Non-tyeable Haemo  
E. coli proliferat  
Haemophilus influe  
Haemophilus influe  
Streptococcus faec  
H. influenzae stra  
Haemophilus influe  
Listeria monocytog  
Lactococcus lactis  
High molecular weight  
Non-tyeable Haemo  
Haemophilus influe  
Haemophilus influe  
Chlamydia psittaci  
C. psittaci protei  
Haemophilus influe  
Haemophilus influe  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Haemophilus influe  
Haemophilus influe  
Arabidopsis thalia  
Arabidopsis thalia  
Eucaalyptus grandis  
ETEC CS4 pilus Csa  
Human kinase, MEK1  
Amino acid sequenc  
Drosophila melanog  
Haemophilus adhesi  
Haemophilus influe



[illegible]

D	b		30	DNDVKATR-----PVLVGGQAVRCVPVPLLTITGPFNIRKRSRGYOGTLVTMLMG	359
O	y		358	SQT 360 : :	
D	b		356	TOT 358	
RESULT 4					
I	d	AAY50814 standard; Protein; 770 AA.			
X	x	AAY50814;			
X	x	AAY50814;			
X	x	AAV50814;			
D	t	17-FEB-2000 (first entry)			
D	e	melanogaster acetyl-choline receptor protein from clone Da7.			
K	w	Acetylcholine receptor; nicotinic; insect; insecticide; screening; neurotransmission; plant protection agent; conductance; AChR.			
P	a	Drosophila melanogaster.			
P	n	DEL98I9829-A1.			
P	d	11-NOV-1999.			
X	x	04-MAY-1998; 98DE-1019829.			
P	r	04-MAY-1998; 98DE-1019829.			
X	p	(FARB ) BAYER AG.			
X	p	Adamczewski M, Oellers N, Schulte T;			
D	r	WPI: 2000-014207/02.			
D	r	N-PADB; AAZ24475.			
P	t	New nucleic acid encoding a nicotinic acetylcholine receptor from insects, used to identify potential insecticides -			
P	s	Example 1a; Page 12-14; 26pp; German.			
C	c	This invention describes a novel nucleic acid (NA) encoding a nicotinic			
C	c	acetylcholine receptor (I) from insects which can be used as an			
C	c	insecticide. Inhibitors of (I) interfere with neurotransmission. (I)			
C	c	(also vectors containing it, its regulatory regions, and antibodies			
C	c	directed against (I)-encoded proteins) are used to screen for: (a) plant			
C	c	protection agents that alter conductance of AChR, potentially useful as			
C	c	insecticides, or (b) genes which encode polypeptides that are involved in			
C	c	formation of functionally related AChR in insects. (I) are also used to			
C	c	isolate and characterize the specified regulatory regions and for			
C	c	recombinant production of (II). This sequence represents an			
C	c	acetylcholine receptor isolated from Drosophila melanogaster.			
X	x				
S	q	Sequence 770 AA:			
Q	y	Query Match          6.3%; Score 119.5; DB 21; Length 770; Best Local Similarity 20.4%; Pred. No. 0.048;			
M	a	Matches 83; Conservative 48; Mismatches 140; Indels 135; Gaps 20			
O	y	9 TLEFFSVLFTFAVSADKLIPGDSTINITEPPDRNESSPKHNIILNN-----54   :           :			
D	b	249 TLFIIFSLYSFLAQLKNSSSSSSSN---SSNSSSTOILLGNKRHSWFLLLYLMLSAK 304   :           :			
D	b	55 --TVAASESHPTTYDNRMTFCGLSSHNTUNGACPTSEBNSSSVSGENITITLOTFEKRLSK 112   :           :			
D	b	305 VCLAGTHEKRLLHD-----LDPTYTKLRPVLNESDPLOJISFGTLTMQIIDVEKKMQLA 359   :           :			
Q	y	113 RELDIIGYKQLLFKSVCNPSSLTNLAHFNCNKKAASGLXYITIPAGEIKLPFGGIWD 172   :           :			
D	b	360 TNVMWLK-----LENMDNLMRANTSDYGVR-----DLR-IIPHRIWK 395   :           :			
O	y	173 ATLKAEVKRRRYSET-----GYTTTINTTIKTIDCKNGIOIWLPQ--FKSDARVDNLRPRTGS 226   :           :			





PR	08-JUL-1992;	92EP-0202080.
PR	14-DEC-1992;	92EP-0203899.

## RESULT 8

DB 396 LYSIDSNITVGPIHITISEVLSDEVEIISKREIASIVVAAPISITIGWIGAM 444

RESULT 8  
AAB01833

ID AAB01833 standard; Protein; 1005 AA.  
XX  
AC AAB01833;  
XX  
DT 11-SEP-2000 (first entry)  
DE  
XX Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.  
XX  
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
KM non-typable Haemophilus influenzae; NTHI; non-encapsulated;  
RV recombinant production; Escherichia coli; antibacterial; vaccine;  
RW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
XX detection; diagnosis.  
XX  
XX Haemophilus influenzae strain K21.  
OS  
PN W020002609-A2.  
13-APR-2000.  
PF 07-OCT-1999; 99WO-CA00938.  
PR 07-OCT-1998; 98US-0167568.  
PR 08-DEC-1998; 98US-0206942.  
XX  
PA (CONN-) CONNAUGHT LAB LTD.  
PI  
PI Loosmore SM, Yang Y, Klein MH;  
DR WPI: 2000-303789/26.  
N-PSDB: AAA52182.  
PT  
PT Nucleic acid molecule for producing recombinant high molecular weight  
PT proteins of Haemophilus which are used as a vaccine to provide  
PT protection against Haemophilus induced diseases in humans -  
XX  
PS Claim 8; Fig 21A-C; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typable (non-encapsulated) H. influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwA genes (AA52175-A52198) and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae strains Joyce, KI, K21, ICDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMWA protein from a non-typable strain of H. influenzae.

Sequence 1005 AA;

Query Match 5.9%; Score 111.5; DB 21; Length 1005;  
Best Local Similarity 20.3%; Pred. No. 0.41;  
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14

60 ESHTLYDMFTPLCLSHNTLNGACFSTSPSSSVSGENITTLQTEKRS-LIKRELDQIK 118  
|::: |::|| |::: |::|| |::: |::|| |::: |::|| |::: |::|| |::: |::||

Db	363	EKNAIISTNNHLITLGGNTVTLGGENSSNNKGINININSKANTVLAQAHGTSHLDKKE----	415
Qy	119	GYKQLLEKSYNCBPGSGLTL--NSAFHCKNKAASGASLYLYIPAGEL-KNLPFGSIW--DA	173
Db	419	--FTLLTLGVNYSVAGNINLIIGTNSNAHIDGNLSIAESAKF-----QCKTNNNLNITGTFPANN	471
Qy	174	TLKLVKRRRSEYGYVTINILIKLMD-----KGNL-----QIWLPOFSDARVD	218
Db	472	TADINIKQGVYKLGQGITVNNGNLNTYTASVNOKTIINGNITNKKGDILNKDIKAAHEIQ	531
Qy	219	L--NLRPITGGGYIGENSDCMFYGYGYSNSSSLEIFRODNNPKSD--GKPYLRKINDYK	275
Db	532	IGGNISQKEBNLTISSDKNTI-----TKRIEDKADTDGQSDSGVASNNALTFIKK	582
Qy	276	ELAYTSLSLLAG--KSLPTNGTSLINADAASLETNNRITAVTMEPISVPLCMGRL	332
Db	583	ELTITLNLNLSIGENFKNAETIAKDNLDLIGKASDSNSNAKQITFDPKVDKSKIS--AGNHNV	640
Qy	333	QIDAKYENPEA-----GQYMNINVTTPSSQTL	361
Db	641	TLNISKVETSNSDGTSGSDNNIGLITISADVTY	675
RESULT 9			
AA001832	ID AAB01832 standard; Protein: 1011 AA.		
AA001832:	AAB01832:		
DT	11-SEP-2000	(first entry)	
DE	Haemophilus influenzae strain K21 HmW2A protein, SEQ ID NO:39.		
KW	HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight; non-typeable Haemophilus influenzae; NTH1; non-encapsulated; recombinant production; Escherichia coli; antibacterial; vaccine; human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis; detection; diagnosis.		
OS	Haemophilus influenzae strain K21.		
PN	WC0200020609-A2.		
PD	13-APR-2000.		
PF	07-OCT-1999; 99WO-CA00938.		
PR	07-OCT-1998; 98US-0167568.		
PR	08-DEC-1998; 98US-0206942.		
PA	(CONN-) CONNAGHT LAB LTD.		
PI	Loosmore SM, Yang Y, Klein MH;		
DR	WPI: 2000-303769/26.		
DR	N-PSDB: AAA52181.		
PT	Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -		
XX	Claim 12; Fig. 21A-O; 307pp; English.		
CC	The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA		

CC proteins. The modified hmwaBC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwa genes (AA52175-A52198)  
 CC and HMWA proteins (AA01824-B01849) from the non-typable H. influenzae  
 CC strains J01, K21, LDC2, PM1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC non-typable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMWA protein from a non-typable strain of  
 CC H. influenzae.

XX Sequence 1011 AA;

Query Match 5.9%; Score 111.5; DB 21; Length 1011;

Best Local Similarity 20.3%; Pred. No. 0.41;

Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

DB 60 ESHTLYDRMTFLCLSHNTLNGACPTSENPSSSSVSGETNITLQTEKRS-LIKRELQIK 118  
 119 GKQKLLFKSVNCSGTL-NSAHFNCNNAAGASLYLTPAGEL-KNIPFGIW-DA 173  
 425 -RTLLGVNVSNGNINILGNSVHIDGNLSIAESAF-OGKTNININITGTETNNG 477  
 174 TKLRVRRYSERYGYTITITIKTD-----KGNL-----QIWLPOKSDARVD 218  
 478 TADINIKQGVKQGITNNGNITNMSVNGKTIINGNITKKKDDINKDKANAELQ 537  
 219 L-NRPFGGTYIGNSVDMCFYDGYSTNSSLERFODNPKSD-GKFLRKINDDT 275  
 538 IGINISQKGNLTISDKNL-----TKRIETKADTDQNSDGSVGSANMLTIKTK 588  
 276 EYATVLTLLAG---KSLPTNGTSLNIDAALENTNRTAVTPEISVPLCWPGR 332  
 589 ELTLDNLNLSGKNAKEITAKDSDLLIGKASDONSNAKQITFDKVKDKIS--AGNHV 646  
 QY 333 OLDKAVENPEA-----GOYMGINVTFTPSQTL 361  
 DB 647 TLNSKVETSDSGSTGSDNNIGLTLISAKDVTY 681

RESULT 10

AA01732

AA01732 standard; Protein; 1529 AA.

AA01732:

XX 25-MAR-2003 (updated)  
 DT 26-APR-1994 (first entry)  
 XX  
 DE High molecular weight protein 4 (HMW4).  
 XX  
 KW HMW; high molecular weight protein; virus; vaccine; influenza;  
 KM epitope; immunity; haemophilus influenzae.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO9319090-A1.  
 XX  
 PD 30-SEP-1993.  
 XX  
 PF 16-MAR-1993; 93WO-US02166.  
 XX  
 PR 16-MAR-1992; 92GB-0005704.  
 XX  
 PA (BARE/) BARENKAMP S J.

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

DR N-PDB; AA049511.

PT High molecular weight surface proteins - of non-typable  
 PT haemophilus which exhibit immunogenic properties

XX Claim 6; Figure 10; 100pp; English.

XX The isolation and purification of the high molecular weight protein  
 CC enables the identification of the major protective epitopes of the  
 CC protein by conventional epitope mapping. These epitopes can then be  
 CC synthesised using standard techniques and incorporated into fully  
 CC synthetic or recombinant vaccines.  
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 1529 AA;

Query Match 5.9%; Score 110.5; DB 14; Length 1529;

Best Local Similarity 20.7%; Pred. No. 0.95;

Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNNHITAYSEHTLYDRMTFLCLSHNTLNGACPTSENPSS-----SVSGETNITL 102  
 744 YNEYSKH--AINSSHNL-----TILGANTLGG-----ENSSSITGNINITNKANVTL 790  
 QY 103 QF-----TEKRSLIKRELQIKGYKQLLFKSVNCPGSLTLNSAHFNCNNAAGASLY 154  
 791 QADTSNNTGLKRRITLGNISVEGNLSLGANANINGNLSI--AEDSTFKGEAS----- 843  
 QY 155 LYTPAGELKNLPRGIIW-DATLKLVRKRYSETGYT-----INTTIKT----- 199  
 844 -----DLNITGFTNNGTANINIKQGVVLTQGITINKGJLNTTNASTGTOKTTING 895  
 QY 200 ----DKGNQIWLPOKSDARVDL-NLRPTGGGTYIGNSVDMCFYDGYSTNSSLER 253  
 896 NINNEKGDNLN-KNIKADAEIOLGNSIKBEKNTLISDKVNI-----TQGITKAG 946  
 QY 254 FQ-----DNPKSGKTYLKRINDTYEIAVTLSSLLAG---KSLPTNGTSLNIDAA 306  
 DB 947 VEGGRSDSSAEANANLTIQ-----TKELKLAGDLNLSGFENKAKETAKNGSDLTIG 1001  
 QY 307 ETMWNRTITANTPEISVPLCWPQ-RLOLDKAVENPEAGGYMGININVT 353  
 DB 1002 NADAKR---VTFDKVKDSKISTDGHVNTLNSEVKTSGSSNAGNDNST 1046

RESULT 11

AA030292

AA030292 standard; Protein; 1601 AA.

AA030292:

XX 14-APR-1998 (first entry)  
 DT  
 XX  
 DE Non-typable Haemophilus high mol.wt. surface protein HMW4.  
 XX  
 KW Non-typable Haemophilus; high molecular weight surface protein;  
 KM HMW4; immunogen; vaccine; otitis media.  
 XX  
 OS Haemophilus influenzae strain 5.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 372  
 FT /note- "encoded by TCR"  
 FT Misc-difference 400  
 FT /note- "encoded by AAT"  
 XX  
 PN WO9736914-A1.

PD 09-OCT-1997.  
XX  
XX 01-APR-1997; 97WO-US04707.  
XX  
XX 01-APR-1996; 96US-0617697.  
XX  
XX (BARE/) BARENKAMP S J.  
XX  
XX Barenkamp SJ;  
XX  
XX WPI; 1997-503038/46.  
DR N-PSDB; AAT90993.  
XX  
XX High molecular weight proteins of non-typable Haemophilus  
PT Influenzae - useful for vaccine production  
XX  
XX  
PS Claim 1; Page 97-102; 183pp; English.

This protein comprises the high molecular weight surface protein HMW4 (123 kDa) of non-typable Haemophilus influenzae strain 5 that has the immunological ability to protect against disease caused by a non-typable Haemophilus strain and is characterised by at least one surface-exposed B-cell epitope that is recognised by monoclonal antibody AD6. The HMW4 amino acid sequence was deduced from an isolated hmw4 gene (see AAT90993). HMW1 (see AAM30293), HMW2 (see AAM30294) and HMW3 (see AAM30291) have also been identified. A conjugate comprising HMW4 linked to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150 amino acids corresponding to at least protective epitope of HMW4 are also claimed. HMW proteins, conjugates and peptides can be used in vaccines, as immunogens for preparation of antibodies and as antigens for detection of these antibodies.

Sequence 1601 AA;

Query Match 5.9%; Score 110.5; DB 18; Length 1601;  
Best Local Similarity 20.7%; Pred. No. 1;  
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNNHTAYSEHTYIDRMFLCLSSHTNLNGACPTSENPSSS-----SYSGETNTTL 102  
DB 816 YNEYSKRH--AINSSHNL-----TILGNTVLGG-----ENSSSITGININITNKANVTL 862  
QY 103 QF-----TERSLIRKREIQKGYKQLFKSVNCPGSLTNSAHFNCKNKAASGASLY 154  
DB 863 QADTNSNMTGLKKRLTLGNTVEGNSLTGNAANIVGNLSL--AEDTFKEAS----- 915  
QY 155 LYIPAGELKNLPFGGIW--DATLKLVRKRRYSERYTYT---INITIKLT----- 199  
DB 916 -----DNLNITGTFTNNGTANINIKOGVYKLOGDIINKGLNITTNASTOKTIING 967  
QY 200 ----DKGNIQIWLPGFKSDARVDL--NLRTGGGTYTIGRNSVDMCFYDGYSTNSSLEIR 253  
DB 968 NITNEKGGDNLN--KNIKADAEIQIGNISQKEGNLTJSSDKVNI-----TNQITTAG 1018  
QY 254 FQ-----DNNPKSDGKFFYLRKINDPKEIAYTSLILAG---KSLPTFGTSLNIDAA 306  
DB 1019 VEGGSDSESEANALTIQ-----TKELKLAGDLNISGFENKRAEITAKKNSDLTIGNASGG 1073  
QY 307 ETNNRITAVTWPETISVPLGMPG--RLQIDAKVENPEAGQYNGNINVT 353  
DB 1074 NADAK-----VTFDKYKDSKISTDGHNTVLNSEVKTSSNGSSNAGNDNST 1118

## RESULT 12

AAB15945  
ID AAB15945 standard; Protein; 2383 AA.

XX AAB15945;  
XX  
XX 05-OCT-2000 (first entry)  
XX  
DE E. coli proliferation associated protein sequence SEQ ID NO:302.

XX  
XX Escherichia coli; E. coli; proliferation; inhibition; screening;  
KW antimicrobial; bacterial growth; antisense therapy; antibacterial.  
XX  
XX Escherichia coli.  
OS  
XX WO200044906-A2.  
XX  
XX  
XX  
XX  
XX 03-AUG-2000.  
XX  
XX  
XX 27-JAN-2000; 2000WO-US02200.  
XX  
XX 27-JAN-1999; 99US-0117405.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
PI Yamamoto RT, Xu HH;  
PI WPI; 2000-514822/46.  
DR N-PSDB; AAA65950.

Novel polynucleotides and polypeptides associated with microorganism proliferation, used to identify inhibitors of bacterial growth and proliferation, for use in antisense therapy -  
Claim 11; Page 224-229; 316pp; English.

AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide sequences derived from Escherichia coli which inhibit E. coli proliferation. AAA65890 to AAA66055 and AAB15986 to AAB16040 represent nucleotide and protein sequences associated with E. coli proliferation. AAA66056 and AAA66057 represent primers used for sequencing E. coli proliferation inhibiting nucleotide inserts in an example from the present invention. Methods from the present invention can be used to identify a proliferation- required gene in a microorganism, by contacting a microorganism with a proliferation- required gene activity inhibitory nucleic acid identified in another organism, and determining if inhibition occurs in the second microorganism. The nucleic acid sequences identified as being required for bacterial growth and proliferation, can be used for antisense therapy for killing bacteria.

Sequence 2383 AA;

Query Match 5.8%; Score 109; DB 21; Length 2383;  
Best Local Similarity 20.0%; Pred. No. 2.5; Mismatches 157; Indels 90; Gaps 15;  
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLTFPAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHTAY---SESHLYDRMT 69  
DB 762 SAKITLASNNGVLANEMANTVSVNVADEGS---NPINHVTFAVLSSGATSFNQNN 818  
QY 70 FLCSSHTNLNGACPTSENPSSSVSGETNTTLQFTKRSLSL-----K 112  
DB 819 ---TAKTDVNGLA--TFDLKSSKQEDNTVEVLENGVQOTLVFVGDSSTAQVDLQKRS 873  
QY 113 RELQIKGYKQI--LFSVNCPSGLTNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIW 171  
DB 874 NEVADGNDVMTATVTDANGNLNDVMTFNVNVAASAKLSQTEVNSHD-----GIA 926  
QY 172 DATLKLVRKRRYSERYTYTITIKTDKGIQIWLPGFKSDARVDLNLRTGGGTYIG 231  
DB 927 TATIL-----SLKNGDYRYTASVSSGQANQVNFIDGSTALITSLV--PSGDITY-- 976  
QY 232 RNSVDMCFYDGYSTNSS---SLEIRFODNNPKSDGKFFYLRKINDPKEIAYTSLILAG 287  
DB 977 -----TNTAPQYMTATLQDKRNGPLKKEITFESVPPD---VASKFSISNGG 1019  
QY 288 KSLPTFGTSL-----NIDAALENNNRITAVTWPETISVPLG 327  
DB 1020 KGMTDSNGVALASLTGTLAGTHMIMARLANSVSDAQPFTVADKRAVVVLQTSKAETI 1079  
QY 328 WPG--RLQIDAKVENPEAGQYNGNINVTFT 355

Db 1080 GNGVDETTLTATVYKDP-SHHPVAGITVNFPT 1108

## RESULT 13

AA01835  
ID AAB01835 standard; Protein; 1095 AA.  
AC AAB01835;  
XX  
DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain LCD2 mature HMMA protein, SEQ ID NO:45.  
XX  
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;  
KM non-typhable Haemophilus influenzae; NTH1; non-encapsulated;  
KM recombinant production; Escherichia coli; antibacterial; vaccine;  
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
XX detection; diagnosis.

OS Haemophilus influenzae strain LCD2.

MO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-303789/26.

N-PSDB; AAA52184.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

Claim 8; Fig 22A-P; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typhable (non-encapsulated) H. influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMMA proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMMA proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMMA. The invention also discloses hmwA genes (AAA52175-52198) and HMMA proteins (AAB01824-B01849) from the non-typhable H. influenzae strains Joyce, K1, K21, LCD2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typhable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMMA protein from a non-typhable strain of H. influenzae.

Sequence 1095 AA;

Query Match 5.7%; Score 107; DB 21; Length 1095;  
Best Local Similarity 24.6%; Pred. No. 1.2;  
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

QY 45 SPKHNILNHTAISHSHLYD-----RMFLCLSSHTINTLNGACPSENSSSS 93  
DB 281 SIRKPIVSN--VHDNHTLFNGNSVVLGGDVNEFFNASSSNHWHGVVISOQNEASE 337  
QY 94 VSGETNITLOFTEKRLSKRELQIKGYKOLPFKSVNCPGLTNSAHCNKNNAASGASL 153  
DB 338 GS-----SLRF-----KSGESTTAF---TIEDTLNATGAGNISLNOVAGIDG 378  
QY 154 VLYPAGELKNLPFGGIWDATLKLRYKRRYSERYTYT---NITIKLTDKGNIGIWP 209  
DB 379 NLOKSLVANKNITFEES--GNITLADKKPIEIKGNITVEKANVTLRGANYGNDKSAI- 434  
QY 210 QEKSDARVLNLRPTGGGTYIGRN-----SYDMCFDGIYSTNSSLEIFQDNNRSD-- 262  
DB 435 SIRGNVTNKGNTLVGSAINIEKNLVEGSAFLANPNYSFVNSGL---FDNOKRSNIS 490  
QY 263 ---GKFLRKINDTKEIAYTIS-----LLAGKSLPTNGTSLNIAD 302  
DB 491 TAKGAHFEDIN-NTKSLNITTTNSDSAYRTIIEG-NITNSNG-DLNIITD 536

## RESULT 14

AA01834  
ID AAB01834 standard; Protein; 1101 AA.

AC AAB01834;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain LCD2 HMMA protein, SEQ ID NO:43.

KM HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KM non-typhable Haemophilus influenzae; NTH1; non-encapsulated;

KM recombinant production; Escherichia coli; antibacterial; vaccine;

KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
XX detection; diagnosis.

OS Haemophilus influenzae strain LCD2.

MO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNAUGHT LAB LTD.

Loosmore SM, Yang Y, Klein MH;

WPI; 2000-303789/26.

N-PSDB; AAA52183.

Nucleic acid molecule for producing recombinant high molecular weight proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans -

Claim 12; Fig 22A-P; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typhable (non-encapsulated) H. influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMMA proteins

CC and the hmwB and hmwC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMW  
 CC proteins. The modified hmwABC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMW. The invention also discloses hmwA genes (AA52175-452198)  
 CC and HMW proteins (AA801824-B01849) from the non-typable H. influenzae  
 CC strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMW protein from a non-typable strain of  
 H. influenzae.

Sequence 1101 AA:

Query Match 5.7%; Score 107; DB 21; Length 1101;

Best Local Similarity 24.6%; Pred. No. 1.2; Mismatches 110; Indels 64; Gaps 15;

Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

QY 45 SPKHNILNNHITAYSESHTLYD-----RMTFCLSSHNTLNGACPTSENPSSSS 93  
 DB 287 SIKRPIVEN---VHDGNHTLFENGANVYLGGDVNFHFNASSNMHTHGVIKSNQNFNASE 343  
 QY 94 VSGENITLQTEFKSLIKRELQIKGYQLLFKSYNCPGSLTNSAHFNCKNAASGASL 153  
 DB 344 GS-----SLRF-----KSEGSTYTAFTTESDLTLNATGNSILNOVAGIDG 384  
 QY 154 YLYIPAGELKNLPFGGIMDATLKLVRKRYSEYGTYTI---NITIKLTGKNIQIOWLP 209  
 DB 385 NLQKSLVANKNKNTFEG---GNITLADKKPTEIKGNITVKEGANTLRANANGNDKSKAL- 440  
 QY 210 QPKSARDLNIKRPFGGGTYIGRN-----SYDMCFYDGYSTNSSSLEIRFQDNKPSD-- 262  
 DB 441 SIRGVTNKGNTLVGSAINIEKRLTVEGSAFNLANPNYSFVNSGIL---FDNQKSKNIS 496  
 QY 263 ---GKFLYLRKINDPTEIAATLS-----LLAGKSLPTNGTSLINAD 302  
 DB 497 IAKGAHFRDIN-NTKSLNITTNSDSAYTITEG-NITNSNG-DLNIITD 542

UUT 15  
 85290

AA85290 standard; Protein; 671 AA.

AA85290;

02-APR-1996 (first entry)

Streptococcus faecalis autolysin.

Lysin; autolysis; culture; lactic acid bacteria; fermentation;  
 cheese; foodstuffs; induction.

Streptococcus faecalis.

W09531561-A1.

23-NOV-1995.

12-MAY-1995; 95MO-NL00170.

12-MAY-1994; 94EP-0201353.

(UNIL ) QUEST INT BV.

Buist G, Kok J, Ledebuer AM, Venema G;

DR WPI; 1996-010946/01.  
 XX Lysis of a culture of lactic acid bacteria in, e.g. cheese  
 PT production - by in situ prodn. of an auto-lysin, regulated by an  
 PT inducible promoter.

PS Disclosure; Page 66-69; 103pp; English.

CC In situ production of a homologous autolysin or a heterologous  
 CC autolysin from a food grade gram positive bacteria, can be used in  
 CC a process for the lysis of a culture of lactic acid bacteria. The  
 CC process can be used in the manufacture of products containing  
 CC cultures of lactic acid bacteria e.g. cheese, where the culture is  
 CC lysed following the completion of fermentation. The enhanced  
 CC induction of the autolysin is performed some hours after the  
 CC fermentation is finished. No extra lysis needs to be added and the  
 CC lysis does not need to be isolated or encapsulated. The time of  
 CC lysis can be precisely controlled. This is the Streptococcus  
 CC faecalis autolysin.

SQ Sequence 671 AA:

Query Match 5.6%; Score 106; DB 17; Length 671;

Best Local Similarity 23.6%; Pred. No. 0.74; Mismatches 103; Indels 94; Gaps 17;

Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESSPKHNIT-LNNHITAYSESHTLYDMTFCLSSHNTLNGACPTSENPSSSV----- 94  
 DB 309 RYATDPYNAKANNVITAY--NLTOYDPPSSGNGTGGGTGVNPGTGSNNQSGTNTTYIVK 366  
 QY 95 SGEI---NITLQF---TEKRSI--IKRELQIKGYQLLFKSYNCPGSLTNSAHFNCKN 146  
 DB 367 SCDTLNKIAAOYGVSANLRSWNGISGDLIFVGOKLIYKKA---SGNTGGSGNGSGNNN 423  
 QY 147 AASGASLYIYPAGELKNLPFGGIMDATLKLVRKRYSEYGTYTIINITIKLTGKNIQI 206  
 DB 424 -QSGINTYTYVKSGLTN-----KIAQIG-----VTV-----ANLRS 455  
 QY 207 WLPQKSDARVDLNI-----RPTGGGTYYGRNSVDMCFYDGYSTNSSSLEIRFQDN 258  
 DB 456 W-----NGISGDLIFVGOKLIYKKGTSGMT-----GSSNGS-----NN 491  
 QY 259 PKSDGKFLYLRKINDPTEIA--YITS-----LLAGKSLPTNGTSLINADA 303  
 DB 492 QSGINTYTYIKSGDPLNKIAAOYGVSANLRSWNGISGDLIFAGOKIIVKKGTSGMTGS 551  
 QY 304 ASLETNMNR 312  
 DB 552 SMGGSNNQ 560

Search completed: August 19, 2003, 09:45:52  
 Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:04:22 ; Search time 3990 Seconds

(without alignments)  
11134.801 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086  
Sequence: 1 atgaataagattattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Imum DB seq length: 0  
Imum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_da:\*
- 2: gb\_hg:\*
- 3: gb\_in:\*
- 4: gb\_cm:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_da:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hg\_hum:\*
- 31: em\_hg\_in:\*
- 32: em\_hg\_other:\*
- 33: em\_hg\_mus:\*
- 34: em\_hg\_pln:\*
- 35: em\_hg\_rod:\*
- 36: em\_hg\_mam:\*
- 37: em\_hg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hgo\_hum:\*
- 40: em\_hgo\_mus:\*
- 41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	AX352333	AX352333 Sequence
2	1086	100.0	7239	AF296132	AF296132 Escherich
3	1086	100.0	7239	AX352351	AX352351 Sequence
4	1086	100.0	7239	AX741418	AX741418 Sequence
5	852.4	78.5	7174	1 ECCOFPA1A	AX741418 Sequence
6	310.8	28.6	5336	6 ECCOCD	M55661 E.coli cfa/
7	310.8	28.6	5336	6 AX741413	AX741413 Sequence
8	309.6	28.5	6956	1 AX216491	AX741413 Sequence
9	298.8	27.5	5798	1 ECCOTABCD	AX741413 Sequence
10	298.8	27.5	5798	6 AX741415	AX741415 Sequence
11	63.4	5.8	1141	6 AX083744	AX083744 Sequence
12	57.6	5.3	61864	3 CEY508BA	AL117200 Caenorhab
13	53.6	4.9	5230	8 AF253047	AF253047 Candida a
14	50.6	4.7	29143	3 AC115594	AC115594 Dictyoste
15	50.6	4.7	88549	3 AC116979	AC116979 Dictyoste
16	50.2	4.6	153477	2 AC006278	AC006278 Dictyoste
17	50.2	4.6	251551	3 AE014844	AE014844 Plasmodiu
18	50	4.6	2000	6 AX655393	AX655393 Sequence
19	49.6	4.6	253924	3 AE014822	AE014822 Plasmodiu
20	49.4	4.5	10279	6 AX279996	AX279996 Sequence
21	49.4	4.5	10279	6 AX346493	AX346493 Sequence
22	49.4	4.5	10279	6 AX356452	AX356452 Sequence
23	49.2	4.5	253151	3 AE014842	AE014842 Plasmodiu
24	48.6	4.5	1141	6 AX083744	AX083744 Sequence
25	48.2	4.4	9504	6 AX323793	AX323793 Sequence
26	48.2	4.4	349980	6 AX344570	AX344570 Sequence
27	48	4.4	3549	3 AF310884	AF310884 Dictyoste
28	47.6	4.4	169794	2 AC004588	AC004588 Plasmodiu
29	47.6	4.4	196149	2 AC004709	AC004709 Plasmodiu
30	47.6	4.4	252650	3 AE014847	AE014847 Plasmodiu
31	47.2	4.3	12940	3 AE001386	AE001386 Plasmodiu
32	46.8	4.3	105682	3 AC116957_3	Continuation (4 of
33	46.6	4.3	110000	2 PFMAL13_10	Continuation (11 o
34	46.6	4.3	110000	2 PFMAL13_05	Continuation (6 of
35	46.6	4.3	110000	2 PFMAL6P1_05	Continuation (6 of
36	46.6	4.3	132254	3 AC116330	AC116330 Dictyoste
37	46.2	4.3	76568	3 MBREV	AF538053 Monosiga
38	46.2	4.3	110000	3 AC116305_1	Continuation (2 of
39	46	4.2	11882	3 AC115612	AC115612 Dictyoste
40	46	4.2	153751	3 AC116551	AC116551 Dictyoste
41	46	4.2	178376	2 BX323051	BX323051 Danio rer
42	46	4.2	236521	2 AC129274	AC129274 Rattus no
43	46	4.2	331039	3 AC116988	AC116988 Dictyoste
44	45.6	4.2	251448	3 AE014819	AE014819 Plasmodiu
45	45.2	4.2	51405	2 AC019512	AC019512 Drosophili

# ALIGNMENTS

RESULT 1	AX352333	1086 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX352333				
DEFINITION	Sequence 9 from Patent WO0181582.				
ACCESSION	AX352333				
VERSION	AX352333.1				GI:18617616
KEYWORDS					
SOURCE					
ORGANISM	Escherichia coli				
REFERENCE	Altboun,Z., Levine,M.M. and Barry,E.M.				
AUTHORS	1				
TITLE	Isolation and characterization of the csa operon (etec-csa p11)				
	and methods of using same				

JOURNAL Patent: WO 0181582-A 9 01-NOV-2001;  
University of Maryland, Baltimore (US)  
Location/Qualifiers

FEATURES  
source  
1..1086  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
1..1086  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAD2854.1"  
/db\_xref="GI:18617617"

CDS  
BASE COUNT 366 a 187 c 202 g 331 t  
ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 1086;  
Best Local Similarity 100.0%; Pred. No. 4.3e-223;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATTAAGATTTATTTATTTTACATTTGTTTCTCTCAGTACTTTTACATTTGCT 60  
DB 1 ATGATTAAGATTTATTTATTTTACATTTGTTTCTCTCAGTACTTTTACATTTGCT 60  
QY 61 GTATCGGAGATTAATTTCCCGGAGATGAAGCATTAATTTTGGCCCGCTGAC 120  
DB 61 GTATCGGAGATTAATTTCCCGGAGATGAAGCATTAATTTTGGCCCGCTGAC 120  
QY 121 AGGAAGCAATCTCTCCCAACATTAATTTAATTAACATTTACAGATAGTAA 180  
DB 121 AGGAAGCAATCTCTCCCAACATTAATTTAATTAACATTTACAGATAGTAA 180  
QY 181 AGTCACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
DB 181 AGTCACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 241 GGAGATGTCACACAGTGAAGATCTGAGTTCATGCGGAGGGAAGCAATTA 300  
DB 241 GGAGATGTCACACAGTGAAGATCTGAGTTCATGCGGAGGGAAGCAATTA 300  
QY 301 ACATTACAAATTTAGGAAAAAGAGTTTAAATAAAGAGCTACAATTAAGGCTAT 360  
DB 301 ACATTACAAATTTAGGAAAAAGAGTTTAAATAAAGAGCTACAATTAAGGCTAT 360  
QY 361 AAACAATTTATTTTCAAAAGTGTAACTGCCATCGGCTTAACACTTAACCTCAT 420  
DB 361 AAACAATTTATTTTCAAAAGTGTAACTGCCATCGGCTTAACACTTAACCTCAT 420  
QY 421 TTAACTGTAATAAAGCGGGCTTCAGGTGCAAGTTTAATTAATTTCCGCTGGC 480  
DB 421 TTAACTGTAATAAAGCGGGCTTCAGGTGCAAGTTTAATTAATTTCCGCTGGC 480  
QY 481 GAACTAAAAATTTGCTTTTGTGTATCTGGATGCTACTGTAAGTTAAGATAAA 540  
DB 481 GAACTAAAAATTTGCTTTTGTGTATCTGGATGCTACTGTAAGTTAAGATAAA 540  
QY 541 AGAGATATAGTGAAGCTATGAAGCTTAACCTTAATTAATTAATTAATTAAT 600  
DB 541 AGAGATATAGTGAAGCTATGAAGCTTAACCTTAATTAATTAATTAATTAAT 600  
QY 601 AAGGAAATATTCAGATATGTTTACTCAGTTCAAAAGTGGCTGGGTGATCTAAC 660  
DB 601 AAGGAAATATTCAGATATGTTTACTCAGTTCAAAAGTGGCTGGGTGATCTAAC 660  
QY 661 TTGGCTCAACTGTTGGGGGACATATATTGGAAGAAATCTGTGATATGCTTTAT 720  
DB 661 TTGGCTCAACTGTTGGGGGACATATATTGGAAGAAATCTGTGATATGCTTTAT 720

DB 661 TTGGCTCAACTGTTGGGGGACATATATTGGAAGAAATCTGTGATATGCTTTAT 720  
QY 721 GATGATATATAGTAAACAGCAGCTCTTTGGAGATTAAGATTTAGATTAACATCTTAA 780  
DB 721 GATGATATATAGTAAACAGCAGCTCTTTGGAGATTAAGATTTAGATTAACATCTTAA 780  
QY 781 TCTGATGGAAATTTTATCTAAGAAAAATTAATGATACACCAAGAAATTCATATAT 840  
DB 781 TCTGATGGAAATTTTATCTAAGAAAAATTAATGATACACCAAGAAATTCATATAT 840  
QY 841 TTGTCATCTCTCTGGGGGGAAGTTTAACCTCAACCAATGGAAGCTATTAATAT 900  
DB 841 TTGTCATCTCTCTGGGGGGAAGTTTAACCTCAACCAATGGAAGCTATTAATAT 900  
QY 901 GCTGACGAGCTTCTCTGGAACAACAACTGGAATTAATTAACAGCTGCACCATCCGAA 960  
DB 901 GCTGACGAGCTTCTCTGGAACAACAACTGGAATTAATTAACAGCTGCACCATCCGAA 960  
QY 961 ATCAGTGTCCGCTGTTGTGTGTGGCTGAGAGCTTTCGAATTTGATGCAAAAGTGAAT 1020  
DB 961 ATCAGTGTCCGCTGTTGTGTGTGGCTGAGAGCTTTCGAATTTGATGCAAAAGTGAAT 1020  
QY 1021 CCCGAGGCTGACATATATAGGGAATTAATTAATTTACTTTTCACACCAAGTACAAACA 1080  
DB 1021 CCCGAGGCTGACATATATAGGGAATTAATTAATTTACTTTTCACACCAAGTACAAACA 1080  
QY 1081 CTCTAG 1086  
DB 1081 CTCTAG 1086

RESULT 2  
AF296132 7239 bp DNA linear BCT 27-FEB-2003  
LOCUS  
DEFINITION  
Escherichia coli csa operon, complete sequence, and insertion  
sequence IS1.  
ACCESSION  
AF296132  
AF296132.1 GI:15419711

KEYWORDS  
Escherichia coli  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 7239)  
Altboum, Z., Levine, M.M., Galen, J.E. and Barry, E.M.  
Genetic Characterization and Immunogenicity of Coli Surface Antigen  
4 from Enterotoxigenic Escherichia coli when It Is Expressed in a  
Shigella Live-Vector Strain  
Infect. Immun. 71 (3), 1352-1360 (2003)

JOURNAL  
MEDLINE  
22483692  
PUBMED  
12595452  
REFERENCE  
2 (bases 1 to 7239)  
Altboum, Z.D., Levine, M.M., Galen, J.E. and Barry, E.M.  
Isolation and characterization of ETEC C54 fimbriae encoding genes,  
and their expression in Shigella flexneri 2a guABA strain CVD 1204  
Unpublished  
3 (bases 1 to 7239)  
Altboum, Z.D., Levine, M.M. and Barry, E.M.  
Direct Submission  
Submitted (14-AUG-2000) Center for Vaccine Development, University  
of Maryland, School of Medicine, 665 W. Baltimore Street,  
Baltimore, MD 21201, USA

FEATURES  
source  
1..7239  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="E1881A"  
/serotype="O25:H42"  
/db\_xref="taxon:562"  
/note="enterotoxigenic"  
283..6849  
/gene="csa operon"  
/note="encodes ETEC C54 fimbrial proteins; CFA/IV"



gene 283. .999  
/gene="csaa"  
CDS 283. .999  
/gene="csaa"  
/note="colonization factor antigen a; may stabilize the major pilin protein"  
/transl\_start=1  
/codon\_start=1  
/product="periplasmic chaperone-like protein CsaA"  
/protein\_id="AAK97134.1"  
/db\_xref="GI:15419712"  
/translation="MKRLFCGLSLITPPVYANANEMITPSKDLKNGSELLRYYSK KEIQIKITTKKINPTEHEHVDMPNDGCVTPPOVYLTPAGSKSIRLQEPHP KKEEIVRYFEAVKPDSEKVNIDNKKLTTELSTNIIAALIRSLPEONSLISNNA RNKIIYNGNVRAGVKDIYFCSSNIDSCVKTINKNIYPERKSEDTLVNNFSYF IKLNHEDIEQGLIQKVP"  
1028. .1531  
/gene="csab"  
1028. .1531  
/gene="csab"  
/note="colonization factor antigen b"  
/codon\_start=1  
/transl\_table=11  
/product="Csa major fimbriae subunit CsaB"  
/protein\_id="AAK97135.1"  
/db\_xref="GI:15419713"  
/translation="MKIKTKIGAMALTFTEVAMASASAVENITVTASVPTDIIDLG GSLLPTAVELTSPASRENTKIAKVTHTNINKRVLVKLVNDPKLTIVLSTKQLP IYVSGCKRLSTADYTFEPALNEFTSGVTGSSODVIAGTAAFPAGNATSGVYS IIMRTLAS"  
1589. .4192  
/gene="csac"  
1589. .4192  
/gene="csac"  
/note="putative outer membrane protein involved in the transport of the major fimbriae subunits across the outer membrane; colonization factor antigen c"  
/codon\_start=1  
/transl\_table=11  
/product="usher protein CsaC"  
/protein\_id="AAK97136.1"  
/db\_xref="GI:15419714"  
/translation="MKKNTLYITITIAMLTPIYSVSGDIPNSFRDLMGODERYEVL YQDTLIRIKITTPHIIKIFYSPESILDKINVKERKLSVLETNFSRGNKSCGN AIIQYNCNIIKTSVDYIVDDVNVNLEIGNEFLDSEAHNDEYHQLSRVKAFTQS QIINVSQSKYKLSVSGNSALGIDTSTAYLVMMANYNKFNCSNNERTINLAFRH DLKRYIYQGRMDRTLDOSISGNFNLPLPDIDGIRITGTOSYIKMTDGIASP VYMTLTPSRVAFRNNOLIGVYIIDSGVNEIDTARLPYCSYDLKIKIPENQYREE IIPNKGSSIGDMQDVFIQSGNITINDKDRITEKONNHSVNAQRLPITKINISYO OGASVINDKNTIEGSLKWNISGLISLNSSEFSLWDNNAKNQOSIYDGFSLSEYH NDKRVNCGRNYNAGWGGYESSASLSIPDLGWTSLGSDYSSEYVKNHILSEYG FYNONIYKGRTOBMOLSTSLKMDYINMPAIGIYNSEBOLDKGGIYSVLTAS RENSINAGSYNSRGKYSNLEFDVGYSTNGIYHEBRMDFNRRNAEERLSGR INNRFGDLNCSPMKNKRNMTNSNHSILGYNSSPALJSDGFWGSSASGLITLAGI IKKSNDRKKNLYKVALGXDLSLSDNDAFLIVPALPILIEDNNTGDNKISVL APTNNDMFLPGNVYVELETKVSVYIGRGEFDKNGTPIPSGAVLNEPHVILDEDEG SEFEYGENEKLFLKGRITTYCOLGRNKHKGIVFDVLCVNSTSSLPDEFVKRPR VQDLAKNDKG"  
4196. .5281  
/gene="csae"  
4196. .5281  
/gene="csae"  
/note="probably found at the fimbriae tip and involved in attachment of the bacteri; colonization factor antigen e"  
/codon\_start=1  
/transl\_table=11  
/product="minor pilin protein CsaE"  
/protein\_id="AAK97137.1"  
/db\_xref="GI:15419715"  
/translation="ANKILFIPTLFFSSVLTPEVAVSADKIPGDESITNINPDRNES SPKHNILNHNITAYSESHLYDMEFTCLNAGACPTSPSSSVSGGTNTL OPEKRSILKRELOIKGYKOLLERKSYNCSGLTLNAGHPNCKNNAASGSILYITAG ELKNLPFGGIWDTLKLRYKRRISRTYTTTITIKLIDKNGIQIIMLPQFKDARYD

LNLKPTGGCTYIGRNSVDMCFYDGYSTNSSLSLEIRPDNNPKSGKFLRKINDTYKE IAYTSLILAGKSLPTNGTSLNIADAALLETWMNRITAVTMEISVPVLCWPGRLQL DAKVENPAGOVYGNINVTETPSSQTL"  
complement(5790. .6119)  
/gene="csad"  
/note="truncated; CsaD"  
/insertion\_seq="IS1"  
repeat\_region 6096. .6870  
gene 6143. .6427  
/gene="isla"  
/note="truncated; Insa"  
6346. .6849  
/gene="islb"  
6346. .6849  
/gene="islb"  
/codon\_start=1  
/transl\_table=11  
/product="Insb"  
/protein\_id="AAK97138.1"  
/db\_xref="GI:15419716"  
/translation="MEGNSPHYGRMFQHDFTSLKRLRPOSTVSRIQPSDVIYCAEMD EOMGYGAKSRORWLFYADVDRLEKTVAVHFGERTMAATGRLMSLSPDVIYKMDG WPIYESRLKGLHVISKRTTORIERNINLROHLARLGRKLSFSKSVELHDVIGHY LNKHYO"  
BASE COUNT 2436 a 1181 c 1345 g 2277 t  
ORIGIN  
Query Match 100.0%; Score 1086; DB 1; Length 7239;  
Best Local Similarity 100.0%; Pred. No. 3.3e-223;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAATAGATTTTATTTATTTTATTTTACATGTTTCTCTCACTACTTTTACATTTGCT 60  
DB 4196 ATGAATAGATTTTATTTATTTTATTTTACATGTTTCTCTCACTACTTTTACATTTGCT 4255  
QY 61 GTATCGGACATATAAATTCGCCGAGATGAAGAATATAATATTTTGGCCCGCTGAC 120  
DB 4256 GTATCGGACATATAAATTCGCCGAGATGAAGAATATAATATTTTGGCCCGCTGAC 4315  
QY 121 AGGAAGAACTCTCCCAACAAATATATTAATTAACATTTTACAGATCACTGAA 180  
DB 4316 AGGAAGAACTCTCCCAACAAATATATTAATTAACATTTTACAGATCACTGAA 4375  
QY 181 AGTCATACCTGTATGATAGATGACTTTTATGTTTGTCTCTCAATACACTTAAT 240  
DB 4376 AGTCATACCTGTATGATAGATGACTTTTATGTTTGTCTCTCAATACACTTAAT 4435  
QY 241 GAGACATGTCCAAACAGTGAATCCTAGAGTTTCATCGTCAAGCGTGAACAAATATA 300  
DB 4436 GAGACATGTCCAAACAGTGAATCCTAGAGTTTCATCGTCAAGCGTGAACAAATATA 4495  
QY 301 ACATTCATATTCGGAAGAAAGAGTTTAAATAAAGAGAGTACAAATTAAGGCTAT 360  
DB 4496 ACATTCATATTCGGAAGAAAGAGTTTAAATAAAGAGAGTACAAATTAAGGCTAT 4555  
QY 361 AAACAATATTTGTTCAAAAGTGTAACTGCCATCCGCGCTTAACACTTACAGTCAT 420  
DB 4556 AAACAATATTTGTTCAAAAGTGTAACTGCCATCCGCGCTTAACACTTACAGTCAT 4615  
QY 421 TTTAAGTGTATTAATAAAGCGGCTTCAAGTGCAGATTTATATTAATATCTGCTGGC 480  
DB 4616 TTTAAGTGTATTAATAAAGCGGCTTCAAGTGCAGATTTATATTAATATCTGCTGGC 4675  
QY 481 GAACTAAAAATTTGCTTTTGTGTGATCTGGAGTCTGCTGAAGTTAAAGTAAAA 540  
DB 4676 GAACTAAAAATTTGCTTTTGTGTGATCTGGAGTCTGCTGAAGTTAAAGTAAAA 4735  
QY 541 AGACGATATAGTGAAGCTATGGAAGTACATATAATATCTACTATTAATTAAGTGAAT 600  
DB 4736 AGACGATATAGTGAAGCTATGGAAGTACATATAATATCTACTATTAATTAAGTGAAT 4795  
QY 601 AAGGGAATATTTAGATATGTTACTGCACTTAAAAAGTACCGCTGCGGCTGCTTAAC 660  
|||||

Db 4796 AAGGAAATATTACAGATATGCTTACCTAGTCAATCAAAAGTGACGCTGCGTGCATTTAAC 4855  
Qy 661 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 720  
Db 4856 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 4915  
Qy 721 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 780  
Db 4916 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 4975  
Qy 791 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 840  
Db 4976 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 5035  
Qy 841 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 900  
Db 5036 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 5095  
Qy 901 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 960  
Db 5096 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 5155  
Qy 961 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 1020  
Db 5156 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 5215  
Qy 1021 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 1080  
Db 5216 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 5275  
Qy 1081 CTTCTAG 1086  
Db 5276 CTTCTAG 5281

RESULT 3  
AX352351 7239 bp DNA linear PAT 06-FEB-2002  
LOCUS Sequence 27 from Patent W00181582.  
DEFINITION AX352351  
ACCESSION AX352351  
VERSION AX352351.1 GI:18617634  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 Alboum, Z., Levine, M.M. and Barry, E.M.  
Isolation and characterization of the csa operon (etec-csa pl1)  
and methods of using same  
Patent: WO 0181582-A 27 01-NOV-2001;  
University of Maryland, Baltimore (US)  
LOCATION/Qualifiers  
1. 7239  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
283. 999  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAD22859.1"  
/db\_xref="GI:18617635"  
/translation="MHKLFCLSLITPPVANANFMIVPISKDKNSSEILIRYSKS  
KEIQIKIYTKKILNPGTBEHEVDMPMDDGFPVTPQKVLIPAGSSSILRQFRIP  
KKEIIRYVFAVRKPSKENVYDNKLTETSEVNIYALIRSLPSEINISINSRNA  
RKNIIYVNGNVRAGVNDYKRCSSNIDSCVKTKHKNITPEKSFILVNNNSYVF  
IKLHEDIEKSGGLQKVP"  
1028. 1531  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAD22860.1"

CDS

Db 4796 AAGGAAATATTACAGATATGCTTACCTAGTCAATCAAAAGTGACGCTGCGTGCATTTAAC 4855  
Qy 661 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 720  
Db 4856 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 4915  
Qy 721 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 780  
Db 4916 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 4975  
Qy 791 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 840  
Db 4976 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 5035  
Qy 841 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 900  
Db 5036 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 5095  
Qy 901 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 960  
Db 5096 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 5155  
Qy 961 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 1020  
Db 5156 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 5215  
Qy 1021 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 1080  
Db 5216 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 5275  
Qy 1081 CTTCTAG 1086  
Db 5276 CTTCTAG 5281

BASE COUNT 2436 a 1181 c 1345 g 227 t  
ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 7239;  
Best Local Similarity 100.0%; Pred. No. 3.3e-223;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAATAAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 60  
4196 ATGAATAAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 4255  
Qy 61 GTATCGGAGATTAATTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 120  
Db 4256 GTATCGGAGATTAATTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 4315  
Qy 121 AGGAACGAATCTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 180  
Db 4316 AGGAACGAATCTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 4375  
Qy 181 AGTACATCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Db 4376 AGTACATCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4435  
Qy 241 GGAGAGTCCCAACAGTGAAGATCCGAGTCAATGCGTCAACGCGTGAACCAATATA 300  
Db 4436 GGAGAGTCCCAACAGTGAAGATCCGAGTCAATGCGTCAACGCGTGAACCAATATA 4495  
Qy 301 ACATTTACATTTACGGAAGAAAGATTATTAAGAAAGAGCTACAAATTAAGGCTAT 360  
Db 4496 ACATTTACATTTACGGAAGAAAGATTATTAAGAAAGAGCTACAAATTAAGGCTAT 4555  
Qy 361 AAACAATTTGTTCAAAAGTGTAACTGCCATCCGCGCTTAACACTTAACGCTCAT 420

Db 4796 AAGGAAATATTACAGATATGCTTACCTAGTCAATCAAAAGTGACGCTGCGTGCATTTAAC 4855  
Qy 661 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 720  
Db 4856 TTGCGTCCAACTGCTGGGGGACACATATATTTGGAGAAATTTCTGTGATATGCTTTTAT 4915  
Qy 721 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 780  
Db 4916 GATGATATGATCTAAGACAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 4975  
Qy 791 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 840  
Db 4976 TGTGATGGAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCT 5035  
Qy 841 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 900  
Db 5036 TTGTACATCTCTTGGCGGATTAAGTTAACTCCACAAATGAGAACTCATTAATAT 5095  
Qy 901 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 960  
Db 5096 GGTGACGAGCTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACATGCGCAGAA 5155  
Qy 961 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 1020  
Db 5156 ATCACTGTTCCGCTGTTGTTGTTGCGTGGAGCTTTCATTAAGATGCAAAAGTGGAAAT 5215  
Qy 1021 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 1080  
Db 5216 CCGGAGCTGACAAATATATGGAATTAATTAATGTTACTTCCACCAAGTAGTCAACA 5275  
Qy 1081 CTTCTAG 1086  
Db 5276 CTTCTAG 5281

BASE COUNT 2436 a 1181 c 1345 g 227 t  
ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 7239;  
Best Local Similarity 100.0%; Pred. No. 3.3e-223;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAATAAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 60  
4196 ATGAATAAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 4255  
Qy 61 GTATCGGAGATTAATTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 120  
Db 4256 GTATCGGAGATTAATTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 4315  
Qy 121 AGGAACGAATCTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 180  
Db 4316 AGGAACGAATCTTCCCGGAGATCAAAAGCATTAATTTTGGCCCGCGTAC 4375  
Qy 181 AGTACATCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
Db 4376 AGTACATCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4435  
Qy 241 GGAGAGTCCCAACAGTGAAGATCCGAGTCAATGCGTCAACGCGTGAACCAATATA 300  
Db 4436 GGAGAGTCCCAACAGTGAAGATCCGAGTCAATGCGTCAACGCGTGAACCAATATA 4495  
Qy 301 ACATTTACATTTACGGAAGAAAGATTATTAAGAAAGAGCTACAAATTAAGGCTAT 360  
Db 4496 ACATTTACATTTACGGAAGAAAGATTATTAAGAAAGAGCTACAAATTAAGGCTAT 4555  
Qy 361 AAACAATTTGTTCAAAAGTGTAACTGCCATCCGCGCTTAACACTTAACGCTCAT 420

Db	4556	AAACATTTTCTTCAAAAAGTTAACTGCCATCCGGCTTAACACTTAATCAGCTCAT	4615	gene	283. .999	/gene="csa"
Qy	421	TTTAACGTGAATAAACGCGGCTTCAGGTGCAAGTTATATATATCTCTGCTGCG	480	CDS	283. .999	/gene="csa"
Db	4616	TTTAACGTGAATAAACGCGGCTTCAGGTGCAAGTTATATATATCTCTGCTGCG	4675			/codon_start=-1
Qy	481	GAACATAAAATTTGCTTTTGGTGTATCTGGAGTCTACTCTGAGTTAAAGTTAA	540			/transl_table=-1
Db	4676	GAACATAAAATTTGCTTTTGGTGTATCTGGAGTCTACTCTGAGTTAAAGTTAA	4735			/protein_id="CAD90914.1"
Qy	541	AACAGATATAGAGACCATGATGAACTTCACTATATAATACATTAATTAATGAT	600			/db_xref="GI:30524216"
Db	4736	AACAGATATAGAGACCATGATGAACTTCACTATATAATACATTAATTAATGAT	4795			/translation="MKKLCGLSLITPEVANANPMIYPIISKDLKNGSELIRVYSKS
Qy	601	AAGGAAATATTCAGATATGTTACTCTCAGTTCAAAGAGAGCTCGCTCATCTTAAC	660	gene	1028. .1531	KEIOYIKIYTKRIINPGTTEHEHVDMPNMDGFEVTPPOKVIIPAGGSKRIETORIP
Db	4796	AAGGAAATATTCAGATATGTTACTCTCAGTTCAAAGAGAGCTCGCTCATCTTAAC	4855	CDS	1028. .1531	KKEIYIRYFEAVKPKDSKENVINDKKLTLESVNIITAIILSLSEQNSISLISNA
Qy	661	TTGCGTCCACATGTTGGGGGACATATATTTGGAAGAAATTCGTTGATATGCTTTTAT	720			IKLNHDEIEKGLIQKVP"
Db	4856	TTGCGTCCACATGTTGGGGGACATATATTTGGAAGAAATTCGTTGATATGCTTTTAT	4915			/gene="csab"
Qy	721	GATGATATAGTACTAACAGACGCTCTTGGAGATTAAGTTTCAGATTAACAATCTTAA	780			/codon_start=-1
Db	4916	GATGATATAGTACTAACAGACGCTCTTGGAGATTAAGTTTCAGATTAACAATCTTAA	4975			/transl_table=-1
Qy	781	TCTGATGGGAATTTTATCTAAGAAATATATGATGACACCAAGAAATTCATATCT	840	gene	1589. .4192	/protein_id="CAD90915.1"
Db	4976	TCTGATGGGAATTTTATCTAAGAAATATATGATGACACCAAGAAATTCATATCT	5035	CDS	1589. .4192	/db_xref="GI:30524217"
Qy	841	TTGTCACATCTCTTGGCGGGTAAAGTTTAACTCCACAACATGGAACGTCAATTAAT	900			/translation="MKIKTIGAMALTFEFAMASAVEKNITVYASVPTIDILQAD
Db	5036	TTGTCACATCTCTTGGCGGGTAAAGTTTAACTCCACAACATGGAACGTCAATTAAT	5095			GSSLPTAEVLITSPASRENNKIKATKHTINVIKKNVAILKYNDRKLTINVDISTOLP
Qy	901	GCTGAGGAGCTCTCTGGAACAAACGGAATAGATATACAGCTGTCACCATGCCAGAA	960			ITVSYGKTLSTADVTPEPAELNFGTSGTVSSQDLVIGATTAQAPTAGNYSQVYS
Db	5096	GCTGAGGAGCTCTCTGGAACAAACGGAATAGATATACAGCTGTCACCATGCCAGAA	5155			ILMTLAS"
Qy	961	ATCAGTGTTCGCGTGTGTGTGGCTGAGCTTTGCAATGGATGCAAAAGTGGAAAT	1020	gene	1589. .4192	/gene="csac"
Db	5156	ATCAGTGTTCGCGTGTGTGTGGCTGAGCTTTGCAATGGATGCAAAAGTGGAAAT	5215			/codon_start=-1
Qy	1021	CCCGAGGCTGGACAATATATGCGTAATATTAATGTTACTTTCACACCAAGTGTCAACA	1080			/transl_table=-1
Db	5216	CCCGAGGCTGGACAATATATGCGTAATATTAATGTTACTTTCACACCAAGTGTCAACA	5275			/protein_id="CAD90917.1"
Qy	1081	CTCTAG 1086				/db_xref="GI:30524219"
Db	5276	CTCTAG 5281				/translation="MKKILFIPTLFSSVLTFFAYSAKDIPGDESTINTIFGRDRNES
LOCUS	AX741418	7239 bp	DNA	linear	PAT 10-MAY-2003	SPKHNLTNNHTIAYSESHTLIDRMATFLCSHNTLNGACPISENPSVSGSETNTTL
DEFINITION	Sequence 8 from Patent WO03022306.					OPTERKSLIKRLOLQIKYKOLLEKSYNCSGLILNSAHNCNKMASSGLVLYIPAG
ACCESSION	AX741418					ELKLNPFEGIMPATLILKRVKRRYSSEYGYTINITIKLDKNGIOMILMPOFSDARVD
VERSION	AX741418.1	GI:30524215				LNMLPFGGTYIGRNSVDKCFYDGYSTNSSLSEIRIODNNPDSCKEFLYRKINDRKE
KEYWORDS						IAYTSLILAGKSIMPTNGTSLINDAASLEENNRITAVTAVPELSVPLICPGRLOL
SOURCE						DAKVENPEAGGYMGNTNVTFFPSSQL"
ORGANISM	Escherichia coli					6346. .6649
REFERENCE	1	Turner, A.K., Greenwood, J., Stephens, J.C., Beavis, J.C. and				/note="unnamed protein product; ISI sequence"
AUTHORS		Darley, M.J.				/codon_start=-1
TITLE	Bacterial Vaccine					/transl_table=-1
JOURNAL	Patent: WO 03022306-A 8 20-MAR-2003;					/protein_id="CAD90918.1"
FEATURES	Acambis Research Limited (GB)					/db_xref="GI:30524220"
source	Location/Qualifiers					/translation="MFGNSPHYGRMPQHDFTSLKLKLRPOSTVTSRIQPGSDVIVCAEMD
	1. 7239					EQMGVGAASRQRLMEYAADRLKRTVAVAFGERMATLGRMAISLSLSPDVIVMTDQ
	/organism="Escherichia coli"					WPLYESRLKGLHVISKRYTORIERHNLTNRHOLRLARLGRKSLSPFSKVELHDKVIGHY
	/mol_type="genomic DNA"					
	/db_xref="taxon:562"					



```
gene /gene="cfa a"
      /product="colonisation factor antigen a"
      /gene="cfa a"
      /gene="cfa b"
      /gene="cfa b"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen b"
      /protein_id="AAC4145.1"
      /db_xref="GI:145509"
      /translation="MKFKTIGAMALTFEAVASAVAEKNTTATASVDPADLLOAD
      GNAPSAAKLAISPASKTEFSTFRVMTQVHTNDATKIVYKLDVFOQLDVLSVQMP
      ISVWGGOVLSTAKEFEAAALGYSAGSVNGVSSQELVISAPKTAGTAPAGNYS
      VVSLVMTLGS"
      /gene="cfa b"
      /gene="cfa b"
      /gene="cfa b"
      /product="colonisation factor antigen b"
      /gene="cfa b"
      /gene="cfa c"
      /gene="cfa c"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen c"
      /protein_id="AAC4146.1"
      /db_xref="GI:145510"
      /translation="MKHKKNLVAVASVALLPYIGVTDIDIDSPFDLMGEDEFEY
      KLYGTLGTHRIKTPHTHIKTPSPESITLKLINLKEKELESVFFNLSRKNNSCQ
      GWTIOYNCNVIKTKSVDIVVDVNVNLFTGNEFLDEAHNDEYHOHSRVKKAQI
      OQOTINLSDSGKRYKRLISGNSALGITDYSVAVLMMNMNKSNGSNNEKTLINSLYF
      RBDILKRYOYGGKMDRTDLSQISGSEFNLPLDIDIGRTGQSYIKNTDKFIA
      SEVTVLTFNFSREAVRNDOLGVWYLDGVELDLAPYGSYDLKLIPEMTOFAS
      EELIPNGRSSISGMOWDIFPOGNIYVNDRYEKKONHSSIMTGRLPRTKIS
      VOQYSVIDNKSTYEGSLKNSGILSGLSNSEFSTLMGNNAKNTOSITYTGFSISE
      YHNDKRVDCGRNNAWMSGCEYSASLIDPLMTSTLAGYSDYSESVYSHLISE
      YGFNONIYKGTORWOLSTSLKMDNDNMPAIGIYNSEBOLDKGQYISVLTFR
      ASRENLNTGYSYNRCNYSNNELEVDGMYSTNGDYGHEAGMRKRNHNAEGLS
      GRINRFGLNGSFNMKNRNTNHNSTHSLTGGYNSSEFALISDGFYWGSGTAGLTKL
      GLIKYSDTKENLYKVTGTLGDSYSLGSDNAFIPVPAVLTASLIEPNNYGNNTIS
      ILAFTNNDFMLPGVYVPEITKVSYSITGKGFDPNGPLSGAHVLPVHLYLDEDG
      GSFPEYTGNEKTLFLKGRITYCOLGKKNVHKGIYFVGVDICDINSTSLPDEFKN
      PRVODLAKNDG"
      /gene="cfa c"
      /gene="cfa c"
      /gene="cfa c"
      /product="colonisation factor antigen c"
      /gene="cfa c"
      /gene="cfa e"
      /gene="cfa e"
      /note="unpublished sequence"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen e"
      /protein_id="AAC4147.1"
      /db_xref="GI:145511"
      /translation="MKKILFITTLPSSFFFPVYSAKNPQSEKNTTIGPDRGGS
      SPYINLNSYLTAYNGSHLIDYRMSFLCISQNTLNGACPSSDAPGTATIDESTNTL
      QETENSLIKRELQIKGYKQPLFKNANCPSLALNSHPCNRQESGATLSLYIPAG
      ELNKLPEFGVNAVLAFLNKRARRDYTYGTYTINIYVNLTKNGNIOIMLPQFSNAVD
      LNLRLPGGTYIGRNSVDKCFYDGYSTNGSSLEIRFODONSXSDGKFLYKIKINDSK
      LYVTLSTLAKNLTPTNGOALNINPASELTWMNRTTAVTMEISVPLVCWGRRLQD
      AKVNPBAQGYKGNKITTPSSQTL"
      /gene="cfa e"
      /gene="cfa e"
      /gene="cfa e"
      /product="colonisation factor antigen e"
      complement(6388..6822)

sig_peptide /gene="cfa a"
mat_peptide /product="colonisation factor antigen a"
gene /gene="cfa a"
      /gene="cfa b"
      /gene="cfa b"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen b"
      /protein_id="AAC4145.1"
      /db_xref="GI:145509"
      /translation="MKFKTIGAMALTFEAVASAVAEKNTTATASVDPADLLOAD
      GNAPSAAKLAISPASKTEFSTFRVMTQVHTNDATKIVYKLDVFOQLDVLSVQMP
      ISVWGGOVLSTAKEFEAAALGYSAGSVNGVSSQELVISAPKTAGTAPAGNYS
      VVSLVMTLGS"
      /gene="cfa b"
      /gene="cfa b"
      /gene="cfa b"
      /product="colonisation factor antigen b"
      /gene="cfa b"
      /gene="cfa c"
      /gene="cfa c"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen c"
      /protein_id="AAC4146.1"
      /db_xref="GI:145510"
      /translation="MKHKKNLVAVASVALLPYIGVTDIDIDSPFDLMGEDEFEY
      KLYGTLGTHRIKTPHTHIKTPSPESITLKLINLKEKELESVFFNLSRKNNSCQ
      GWTIOYNCNVIKTKSVDIVVDVNVNLFTGNEFLDEAHNDEYHOHSRVKKAQI
      OQOTINLSDSGKRYKRLISGNSALGITDYSVAVLMMNMNKSNGSNNEKTLINSLYF
      RBDILKRYOYGGKMDRTDLSQISGSEFNLPLDIDIGRTGQSYIKNTDKFIA
      SEVTVLTFNFSREAVRNDOLGVWYLDGVELDLAPYGSYDLKLIPEMTOFAS
      EELIPNGRSSISGMOWDIFPOGNIYVNDRYEKKONHSSIMTGRLPRTKIS
      VOQYSVIDNKSTYEGSLKNSGILSGLSNSEFSTLMGNNAKNTOSITYTGFSISE
      YHNDKRVDCGRNNAWMSGCEYSASLIDPLMTSTLAGYSDYSESVYSHLISE
      YGFNONIYKGTORWOLSTSLKMDNDNMPAIGIYNSEBOLDKGQYISVLTFR
      ASRENLNTGYSYNRCNYSNNELEVDGMYSTNGDYGHEAGMRKRNHNAEGLS
      GRINRFGLNGSFNMKNRNTNHNSTHSLTGGYNSSEFALISDGFYWGSGTAGLTKL
      GLIKYSDTKENLYKVTGTLGDSYSLGSDNAFIPVPAVLTASLIEPNNYGNNTIS
      ILAFTNNDFMLPGVYVPEITKVSYSITGKGFDPNGPLSGAHVLPVHLYLDEDG
      GSFPEYTGNEKTLFLKGRITYCOLGKKNVHKGIYFVGVDICDINSTSLPDEFKN
      PRVODLAKNDG"
      /gene="cfa c"
      /gene="cfa c"
      /gene="cfa c"
      /product="colonisation factor antigen c"
      /gene="cfa c"
      /gene="cfa e"
      /gene="cfa e"
      /note="unpublished sequence"
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen e"
      /protein_id="AAC4147.1"
      /db_xref="GI:145511"
      /translation="MKKILFITTLPSSFFFPVYSAKNPQSEKNTTIGPDRGGS
      SPYINLNSYLTAYNGSHLIDYRMSFLCISQNTLNGACPSSDAPGTATIDESTNTL
      QETENSLIKRELQIKGYKQPLFKNANCPSLALNSHPCNRQESGATLSLYIPAG
      ELNKLPEFGVNAVLAFLNKRARRDYTYGTYTINIYVNLTKNGNIOIMLPQFSNAVD
      LNLRLPGGTYIGRNSVDKCFYDGYSTNGSSLEIRFODONSXSDGKFLYKIKINDSK
      LYVTLSTLAKNLTPTNGOALNINPASELTWMNRTTAVTMEISVPLVCWGRRLQD
      AKVNPBAQGYKGNKITTPSSQTL"
      /gene="cfa e"
      /gene="cfa e"
      /gene="cfa e"
      /product="colonisation factor antigen e"
      complement(6388..6822)
```

```
CDS
      /gene="cfa d"
      complement(6388..6822)
      /gene="cfa d"
      /note="truncated cfa-d-like protein with a stop codon and
      frameshift deletions between bp 6029 and 6824."
      /codon_start=1
      /transl_table=11
      /product="colonisation factor antigen d"
      /protein_id="AAC4148.1"
      /db_xref="GI:145512"
      /translation="MDFKTEEEKMKIKINIMIHITYVITSNCIMDIYSEEEKITCS
      SNLVLENGVINSVRIQKQILSEKRYLFRNLGDLRLKNAALMITGMSKIDINDC
      RNSKRIITTEVDKTLIDVLKKNINSYDVSFFISLYLSIKI"
BASE COUNT 2441 a 1129 c 1271 g 2332 t 1 others
ORIGIN

Query Match 78.5%; Score 852.4; DB 1; Length 7174;
Best Local Similarity 87.2%; Pred. No. 4.9e-173;
Matches 947; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

1 ATGAATTAAGATTTTATTTATTTTACATTTGTTTCTCTCACTTTTACATTTGCT 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
4783 ATGAATTAATTTTATTTATTTTACATTTGTTTCTCTCACTTTTACATTTGCT 4842
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
61 GATTCGCAATTAATTTCCCGAGATGAAGCATTAATTTTGGCCGCGTGC 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
4843 GATTCGCAATTAATTTCCCGAGATGAAGCATTAATTTTGGCCGCGTGC 4902
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
121 AGAAGCAATTTTCCCGCAACATTAATTTAATTAACCATTTTACAGTACATGA 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
4903 AGGAGGGAATCTTCCCGCAATTAATTTAATTTAATTTTCACTTTACACATCA 4962
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
181 AGTCAACTCTGATGATAGATGACATTTTATTTTATTTTCTCTCACTTTTAC 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
4963 AGCCATCACTGATGATAGATGACATTTTATTTTATTTTCTCTCACTTTTAC 5022
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5023 GGAGCATGCCAAGCATGATGCCCTGCACGCTACAAATGATGGCCAAATATA 5082
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 ACATTCAATTAAGGAAAAAGAAAGTAAATTAAGAGAGCTAACAAATTAAGCTAT 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5083 ACATTCAATTAAGGAAAAAGAAAGTAAATTAAGAGAGCTAACAAATTAAGCTAT 5142
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
361 AAACAATTTTGTTCAAAGTGTAACTGCCATCCGCGCTAACACTTAACAGCTCAT 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5143 AAACAATTTTGTTCAAAGTGTAACTGCCATCCGCGCTAACACTTAACAGCTCAT 5202
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
421 TTTAAGTGAATTAAGGCGGCTTCAAGTGAAGTTTATTTATTAATTTCTGCTGC 480
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5203 TTTCATGTAATTAAGGAAACAGCTTCAAGGCTTATTCGTTATACATTAACAGCTGC 5262
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
481 GAACATAAATTTGCTTTTGGTGTGATCTGGAGCTACCTGTAAGTAAAGTAA 540
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5263 GAATTAATTAATTTGCTTTTGGGAGGCTGGAAGCGCTTGAAGCTAAATTAAGTAA 5322
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
541 AGACGATATAGTGAAGCTATGAGACTTACATTAATTAATTAATTAATTAAGTAA 600
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5323 AGACGATATAGTGAAGCTATGAGACTTACATTAATTAATTAATTAATTAAGTAA 5382
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
601 AAGGAAATATTCAGATATGTTTCACTCAAGTCAAAAGTGACGCTGCGTCAATCAAC 660
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5383 AAGGGAATATTCAGATATGTTTCACTCAAGTCAAAAGTGACGCTGCGTCAATCAAC 5442
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
661 TTGCGTCAACAGTGTGTGTATATTCGGAAGAAATTCGTATATATGTCCTTTAT 720
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5443 TTGCGTCAACAGTGTGTGTATATTCGGAAGAAATTCGTATATATGTCCTTTAT 5502
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
721 GATGATATATGATTAAGACAGCTTTTGGAGATTAAGTTTCAGATTAACATCTTAA 780
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
5503 GATGATATATGATTAAGACAGCTTTTGGAGATTAAGTTTCAGATTAACATCTTAA 5562
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
781 TCTGATGGAAATTTTATCTTAAGGAAATTAATGATGACACCAAGAAATTCATTA 840
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

```

Db      5563 TGTGATGAAATTTTATCTAAGAAATTAATGATGACTCCAAAGAACTGTATACACT 5622
Oy      841 TTGTCACTTCTCTTGGCGGGTAAAGTTTAACCTCCACAAATGAAAGCTATAATAT 900
Db      5623 TTGTCACTTCTCTTGGCGGGTAAAGTTTAACCTCCACAAATGAAAGCTATAATAT 5681
Oy      901 GCTGACGAGCTTCTCTGGAACAACAACGGAATAGCAATTAACGCTGTCACCATGCCAGAA 960
Db      5682 --TAACTACTGCTTCTCTGGAACAACAACGGAATAGCAATTAACGCTGTCACCATGCCAGAA 5739
Oy      961 ATCACTGTTCCGGTGTGTGTTGGCTGGAGCTTGCAATGATGCAAAAGTGGAAAT 1020
Db      5740 ATCACTGTTCCGGTGTGTGTTGGCTGGAGCTTGCAATGATGCAAAAGTGGAAAT 5799
Oy      1021 CCCGAGGCTGCAATATATGCGTAAATATTAATGTTACTTTCACACCAAGTGTCAACA 1080
Db      5800 CCCGAGGCTGCAATATATGCGTAAATATTAATGTTACTTTCACACCAAGTGTCAACA 5859
Oy      1081 CTCTAG 1086
Db      5860 CTCTAG 5865

```

```

RESULT 6
LOCUS   ECCOCCD 5336 bp DNA linear BCT 07-JUL-2002
DEFINITION E.coli cooc and cood genes.
ACCESSION X76908
VERSION X76908.1 GI:488735
KEYWORDS cooc gene; cood gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

```

```

REFERENCE
AUTHORS Froehlich,B.J., Karakashian,A., Melsen,L.R., Wakefield,J.C. and Scott,J.R.
JOURNAL Cooc and Cood are required for assembly of CSI p111
MEDLINE Mol. Microbiol. 12 (3), 387-401 (1994)
PUBMED 94344028
7915003
2 (bases 1 to 5336)
Scott,J.J.
Direct Submission
JOURNAL Submitted (24-DEC-1993) J.J. Scott, Emory Uni. School of Medicine,
AUTHORS Dep. of Microbiol. and Immunol., 1510 Clifton Rd, Atlanta 30322,
Georgia, USA
COMMENT Entry cited as #76908 in the journal.
FEATURES
Location/Qualifiers
source 1..5336
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="LMC10"
/db_xref="taxon:562"
/clone="pEU494"
76..79
RBS /db_xref="GI:488736"
gene /db_xref="SPTREMBL:Q47114"
CDS /translation="MIGKSSKVVIVLSVIGSSGPAKXNVLVDIPESFRLMGEOD
ELFELVRLGOSLGVHRIKSPPTVAFAESPDLIDKIEINKREADLRILMGSEFQRMG
NMSCGYTGNNCNVYIKNTVAIVLVDVENTLNFIGNELFASSENSDYQPSKNTK

```

```

17
RBS
gene
CDS

```

```

KAFHSGTINSDTGNFENLSITVGSLGTDNSYAILGMAANTNRRKSYNNEOSIN
SLYRHDENKFFYQLGRIDRSDLSQSGGNFNDLLPVPDIYOMRAGTOSYIKNG
KSVASPTVIMLTHFSRVAEYRNQQLGIVEDAGISELDEPRLPDGNILKLFEEG
OLVREIYVPMKSSSISGDTMDYVVGAGDIIINDNGRVEKOKHKSAINSLGLPFL
RNIAVVGGAVIDKNNYERGIILNNSGILDSLSKSTFELRGDDTHGNOYVSTDFP
SLSTYHNDKRDDCGKDYNMWMSGCEYSNLSLSPKGNMSTLAYSITSTYTRID
AVSERPYTYIKGRTRKWLASTVVRKGDNIINPTLGINSEOKOMADKGYLSTL
TRVDGKSLNAGYSYNSRQYTSNDLVECHLVDNVSYSRELARSNGRYTEGG
VSGRINNRFGDNLGTLNANKRSHDPTSLAGYSSSFALTGDIYWGSGSLTNL
SGCATRVKSNDESESLNVKSSGSHSLGSDSLFIPVPLMQASLIEENTKSKN
IDVLAPTKNEFPM.PGSYPTIDVSANYSFYVGVGVYKGRPGSGATILANONVLE
NGSGPSESENEKELFLKDKTIIYCSLDRKEMNGIAFGEVACNSTIKELPEKLY
TNSRIHDLAYNODE"
85..156
/gene="cooc"
/citation=[1]
157..2700
/mat_peptide
/gene="cooc"
/product="Cooc"
/citation=[1]
2689..2692
RBS /gene="cooc"
/citation=[1]
2700..3791
gene /gene="cood"
2700..3791
CDS /gene="cood"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cooc"
/protein_id="CA54230.1"
/db_xref="GI:488737"
/db_xref="SPTREMBL:Q47115"
/translation="MKRIFELSIIFSAYVSAGRYPEPTYGNLTKSPQADRLDSNOS
PIYVIFNHNVAYSLSISYDRIVFICTSSPNVNGACPRITGSGYGYCTTTLQGT
EKSLIKRINILANKRPINENQCDNSMLKAVLSKWSGCAHGNAGTILNLTIPAG
EINKLPREGIWEATLILRLSKRYGVSSSTHGNVYNTVDTLTDGNTQVWLPGFHSNP
RVLDNLRLPILGNKYSGNSLDMCFDYDGYSTNSDSMVIKFDODNPTNSSEYVNLKIGGT
EKLPYASLMLGKERIFYPVNGOSFTINDSVILETNNMRYTAVAMPVNVPLCWPARL
LNLADVNPADGQYSGOYITFTPSVENL"
2700..2753
sig_peptide
/gene="cood"
/citation=[1]
2754..3788
mat_peptide
/gene="cood"
/product="Cooc"
/citation=[1]
4476..5536
repeat_region
/note="bases 1-887 of IS2"
/insertion_seq="IS2"
BASE COUNT 1553 a 1021 c 1206 g 1556 t
ORIGIN
Query Match 28.6%; Score 310.8; DB 1; Length 5336;
Best Local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;
Oy 7 AAGATTTTATTTTATTTTATTTTCTCTCAGACTTTTATCATTTGCTGTTCG 66
Db 2703 AAAAAATATTTATTTTCTCATCATTTTCTCGGTTGTCAGTGCCGGCGCATAC 2762
Oy 67 GCAGATAAAATTCGCGAGTGAAGCATATATATTTTGGCCCGGTGACAGGAAC 126
Db 2763 CCGGAACATACAGTAGGATATCTGACGAAGGTTTCAAGCCCTCGTGGATAGAAC 2822
Oy 127 GAATCTTCCCGCAACAATATATTTAATATACATATTACAGATCAGTGAAGTCAT 186
Db 2823 GTACAATCCACCAAAATATATACATCTTTACGATATCTGTGATGATATAGTTCAT 2882
Oy 187 ACTGTGATGATGAGATGACTTTTATGTTGTTCTCTCACATATACATTAATGAGCA 246

```

```

Db      2883  AGCTATATGACAGATTTGTTTTTATGCTACATCTCTCCAAATCCGTTAATGCTGCT 2942
QY      247  TGTCCAAACGAGTGAAGATCTAGACTGATCGGTGACGGTGAACAAATATACATTA 306
Db      2943  TCCCCAACCATTTGGAACATCTGAGTTCAT-----ACGGATCTCAACCAATACCTTG 2996
QY      307  CAATTACGGAAGAAAGATTTAATAAAGAGAGCTCAATTAAGGCTATTAACA 366
Db      2997  CAGTTTACGAAAAAGAGTGTGATTAAGAAATTAATTAATCTGACGTAATAAGAA 3056
QY      367  TTAATGTTCAAAAGTTTAAGTCCATCCGCTTACACTT-----AAGTCAGTCAAT 420
Db      3057  CCAATATGGAGATACAGATTGCGACTTTAGCAATCTAATGCTGTAATTCGAATG 3116
QY      421  TTTAATCTGTAATTAACG---CGGCTTACGTCGTAAGTTTATTAATTAATCTCTG 477
Db      3117  TGGAGCTGTGGGGCTCAGGAATGCTAACGGAACTCTTAATCTGATATACCTGTA 3176
QY      478  GCGCACTAAAAAATTTGCTTTGCTGATCTGGGATGCTACTGTAAGTTAAGATA 537
Db      3177  GGAGAAATCAACAAATGCTTTTGGAGGATATGGGCAAACTCTGATCTACGCTTA 3236
QY      538  AAAAG-----ACGATATAGTAGACCTTAAAGAACTTAACATTAATACATAT 588
Db      3237  TCAAGATATGGGAACTCAGTACACCCATTACGGCAATTAATACGTAATATTCG 3296
QY      589  AAATTAAGTAAAGGAAATATTCAGATATGCTTACCTCAGTTCAAAAGTACGCTGC 648
Db      3297  GATTTAAGTAAAGGATTAATTCAGTATGCTTCCAGGCTTCCAGCAACCGCGCT 3356
QY      649  GTGCAATCTAAGTCTGCTCAACTGCTGGGGGACATATTTGGAAGAAATTCGTGAT 708
Db      3357  GTAGACCTGAATCTGGCCCTATCGGTAATTAATTAATAGTGAATTCACATCAG 3416
QY      709  ATGCTTTTATGATGATATAGTACTAACACAGCTTTTGGAGATTAAGATTTAGAT 768
Db      3417  ATGCTTTTATGATGATATAGTACTAACACAGTATGATATTAAGTTCCAGGAT 3476
QY      769  AACAATCTAATTAATCTGATGGAATTTATCTAAGGAAATTAATGATGACACCA 828
Db      3477  GATTAATCTAATTAATCTGATGGAATTTATTAATTAATTAATTAATTAATTA 3533
QY      829  ATTTGATATATCTTTGCACTTCTCTGGGGGTAATTAATTAATTAATTAATTA 888
Db      3534  TTACATATGCTGCTTCACTGCTTATGGAAGAAATTTATTAATTAATTAATTA 3593
QY      889  TCATTAATATTTGCTGACAGCTTCTCTGGAACAACTGGAATTAATTAATTA 948
Db      3594  TCATTAATATCAATGACAGTATGCTGCAAACTGGAATTAATTAATTAATTA 3653
QY      949  ACCATGCCAGAAATCAGTGTCCGGTGTGTTGCTGGAAGCTTTCGAATTTGA 1008
Db      3654  GGTATCCGGAAGTAAATTTCCAGATTAATGCTGCGCAGCAAGATTTGATTA 3713
QY      1009  AAAGTGAAGAAATCCGAGGCTGACATTAATGGAATTAATTAATTAATTAATTA 1068
Db      3714  GATGTAAGATCCCGATGACAGAGTATTAAGAGATTAATTAATTAATTAATTA 3773
QY      1069  AGTAGTCAAA 1078
Db      3774  AGTGTGAAA 3783

```

```

RESULT 7
AX741413 5336 bp DNA linear PART 10-MAY-2003
LOCUS AX741413
DEFINITION Sequence 3 from Patent WO03022306.
ACCESSION AX741413
VERSION AX741413.1 GI:30524203
KEYWORDS
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

REFERENCE 1
AUTHORS Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and
Darsley,M.J.
TITLE Bacterial vaccine
JOURNAL Patent: WO 03022306-A 3 20-MAR-2003;
Acambis Research Limited (GB)
FEATURES
source
    /organism="Escherichia coli"
    /mol_type="genomic DNA"
    /db_xref="taxon:562"
    85..2703
    /note="unnamed protein product; cooc"
    /codon_start=1
    /transl_table=11
    /protein_id="CAD90906.1"
    /db_xref="GI:30524204"
    /translation="MIGKSSRVAVIVLSYIGSSGFAKRYNLVDIPESFRLMGEOD
    ELLEVRVLSQSLVHRKSTPTTVAEFPEDNLDRLEIKRGKREADLRVLMRSGFORNG
    NMSCGYTGQNNCNTIKTNTVAIVDVDEVNVLNFTGNFELASGENDSDYQPSKNTK
    KAFHSQRTINLSDTGNVENLSTVGSGISGITPNSYAILGMAANYNRYKSYANTNEDSIN
    SLVFRHDEPEKNRYVYOLGRIDRSDLSQSSGGMNPDLPPVDTYGRACGTQSYIKNTG
    KSYASPTVIMLHRSREAYRNGQLGVTYLDAGISELTERLPDNTDLKLTFOE
    QVREIVEFNSGSSIGDTHNDVFOADDIINDGRIVEKKNRKSAINSLRLPLT
    RNLAVOLGAVIDNKNRYETGLMNSGLDGLNSKFTFLFDGDDHGNQYNSYTDG
    SLSPYHNDKRVLDGCKDYMNGSGCYESASLSIPVKMNSTLAYSNYSYVRYD
    AVSEVYVYVYKGRKRMQLTASTVRYMDYNIIMPLIGVNSEOKOMADKAGYLSLT
    TVDQDGLSNAGYSYNSRGNTSNDAPVEGHLVDITNYSRELARSVGNRYTEGG
    VSGRIINNRGDIINGLNKRSKSHDTHSLAGYSSSPALTTDITNGSGASGTLN
    SSGIYRVKSNDEBELNAYKGSYGHYSLSGSDSLFIPPALMQSLTIEENTNSKN
    IDVLAPTKTFEFLMPSVYPIDVASNVSTYVGRVAKGRPLSGAVITLMAQNTLDE
    NNGFSESENEKELEFLDKDTIYCSIDRSRMRNGIAFVEGVAACNSTYIKELPEKLV
    TNSRIHDLAYNQDTE"
    2700..3791
    /note="unnamed protein product; cood"
    /codon_start=1
    /transl_table=11
    /protein_id="CAD90907.1"
    /db_xref="GI:30524205"
    /translation="MKKIFPLSIFSAVNSAGRYPETVGNLTKSPQAPRLDRSVOS
    PLYNFTNNAVYSLSLSLSDRIYFLCTSSSPVNGACCTTCTSGVQGTITTTQFT
    EKRSLIKRININAGNKPPIWENQSCDSFSLNLSKMSWCGAHGANGLNLLYIPAG
    ELNKLPEFGIWEATLILRLSRGEVSTHYVNTVITVDLKDKNINQWLPGEFSNP
    RYDLNLRPIGNKRYSGNSLDMCFYDGTSTNDSMSYIKQODNPNNSSSYNIYKIGGT
    EKLPRVSLIMEKEIFYPVNGOSFTINDSVLETNMNRVTAVAMBEVNPVLCWPARL
    LNDVADNADAOYSGQIYITFTPEVENL"
BASE COUNT 1553 a 1021 c 1206 g 1556 t
ORIGIN
Query Match 28.6%; Score 310.8; DB 6; Length 5336;
Best local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;

```



LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
307	CAATTACGAAAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTATAACAA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
2997	CAGTTTACAGAAAAAGAGAGCTGATATAAAGAAATATATCTTGCGATATATAGAA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
367	TTATTTGTCAAAGCTGTTAACTGCCCATCCGCCCTAACACTT-----ACTCAGCTCAT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3057	CCAAATATGGGGAAATCAGAGAGTTGGAGACTTTAGCAATCTAAAGTGTTGAATTGCAAGTCT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
421	TTTAACTGATATAAAAAAG---CGGCTTCAGGTCAAGTTTATATATATTCCTGCT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3117	TGAGAGCTGTGGGGCTCAGAGAAATGCTAACGGAACACTCTTAAATCTGTATATCCCTCA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
478	GGCGAACTAAAAAATTTGCCCTTTTGGTGTATCTGGAGTGCCTACTGAAATTAAGACTA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3177	GGAGAAATTCACAAATTTGCCCTTTTGGAGGGATATGGAGCAACTGATCTATTCAGCTTA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
538	AAAG-----ACGATATAGTAGAGACCTATGAACTTACACTATATAATCTACTTT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3237	TCAAGATATGGCGAAGTCAGTAGACACCCATTACGGCAATTTATCCGTAATAATTAACGGTT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
589	AAATTAATCGATTAAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTAGAGCTGCG	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3297	GATTTAACTGATTAAGGATATATTCAGGATATGCGTTCCAGGGTTTCACAGAACCCGGT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
649	GTCGATCTTAACCTGCTCCCAACTGCTGGGGGACATATATTTGGAAGAAATCTGTTGAT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3357	GTAAACCTCGAATCTGCGCCCTATCGGTATTAATATATAGTGTAGTAATTAACCTGCAC	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
709	ATGTCCTTTATGATGATATAGTATAGTATACACAGACGCTTTTGGAGATTAAGATTCAGGAT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3417	ATGCTTTCTATGATGATATAGTATAGTATACACAGCTATAGATGATTAATTAAGTCCAGAT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
769	AACAATCTTAATCTGATGGGAAATTTATCTAAGAAATTAATATGATGACACCAGAA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3477	GATATATCTTCAATTCATCTGATATTAATATCTTTATTAAGATA---GGGGCAGCTGAAAA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
829	ATTGCAATATATTTGTGCTCTCTTG6GGGTTAAAGTTAACTCCACAAATGAGAACG	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3534	TTACCAATATGCTGTTTCTACCTGCTTATGAGGAGAAAAAATATTTATCCAGTAAATGCTCA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
889	TCATTAATATATTTGTGACGACAGCTTCTGTGAAACAAACTGGAATAGATTAACAGCTGC	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3594	TCATTTACTATCAATACAGATAGTATGACTGAAACAAACTGGAATCGAGTAAACGCACTT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
949	ACCATGCCAGAAATCAGTGTCCGCTGTGCTGTGCTGACGCTTGCAATTTGGATGCA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3654	GCATATGCCGGAAGTAAATGTTCCAGTATTAATGCTGGCCAGCAAAATTCCTATTAATAGCT	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
1009	AAATGGAAAAATCCGAGAGCTGACCAATATATATTTATTTATTTACTCTTCAACCA	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3714	GATTAATATGCTCCCGATCGACAGACAGATATCAGAGACAGATATATTAACCTTTACACCC	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
1069	AGTAGTCAAA 1078	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			
3774	AGTGTGAAA 3783	AY216491	6956 bp	DNA	linear	BCT 12-MAR-2003			

JOURNAL Unpublished  
 2 (bases 1 to 6956)  
 REFERENCE Biediger, W. and Wolf, M.K.  
 AUTHORS Direct Submission  
 TITLE Submitted (09-JAN-2003) Enteric Infections, Walter Reed Army  
 JOURNAL Institute of Research, 503 Robert Grant Ave, Silver Spring, MD  
 20910-7500, USA  
 FEATURES  
 source location/Qualifiers  
 1. 6956  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /strain="DS37-4"  
 /db\_xref="taxon:562"  
 /plasmid="unnamed"  
 /note="naturally-occurring plasmid"  
 166. 879  
 /gene="csbB"  
 166. 879  
 /gene="csbB"  
 /note="CS17 fimbrial accessory protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CsbB"  
 /protein\_id="AA060095.1"  
 /db\_xref="GI:28932775"  
 /translation="MKRLFLSLMLPEPTKAMFYPVKAKMYPISKIKGASSELLIRISKSD  
 TOYIKYAFKRVGPTGKDEXEIDAPNMEGALVYTPAVIILPSSGSKSVRLTQUNAPPT  
 EEEYRVFEEVSKPEAOENSSGNKTKLIDLVNVIYALILRILPINKTASVISPSFN  
 NLIKNTNGVRLGIMVFECKSSINDCYKSKGNKVVGYLLDPMIYNDGYSHLFL  
 DTKDESSKTDNLIDYPIIT"  
 933. 1439  
 /gene="csbA"  
 933. 1439  
 /gene="csbA"  
 /note="CS17 fimbrial subunit; structural subunit of CS17  
 pilus"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CsbA"  
 /protein\_id="AA060096.1"  
 /db\_xref="GI:28932776"  
 /translation="MKRLKIGAMAAATLEPATMAASAVEKNTYRPASVDEPKLLDAD  
 GTSLPSIALITYSSASNNFEVSLNTHINDSKGVYKLSAPVLSNIMKRNQSP  
 MKYTLGKTLNTIDTEPYDTLNFGISGVENSVTQDLTHADTQGTAPAGNYGII  
 SLIMTKRT"  
 1506. 4124  
 /gene="csbC"  
 1506. 4124  
 /gene="csbC"  
 /note="putative fimbrial usher"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CsbC"  
 /protein\_id="AA060097.1"  
 /db\_xref="GI:28932777"  
 /translation="WYSKNSKSLIYLFELSSPLGPAOKNKTITNPESPRLMGEDD  
 GLEFVEYFGSGIVGHRIKLKLPTTVAEPSPSLDKLEINKERADRYLMRGSPQRNG  
 NMSGCGTGNQNCNTTKNTNVAIVDDVENLVNLFIGNERLASGENDSDIYQPSKNTK  
 KAIHQSTLINSDPTGYNENLSTIVGSLGITDQNSVAILGMAANYNRYKSKYNEOSIN  
 SLVFRDPEKNEYQDGRIDRDLSSGSGENFEDLLPVDIYGMAGTQSYIKKTG  
 KYSAPVTIMLTHFSRVEAYRNGQLLGWVLDAGISELDELRDPDNGYILKEINTKSN  
 OLAREELIYFPNKSISGIDTHMDVFOAGQIINDNRYEYKOKNHSKAINSGRLPLT  
 RNLAOLGGAVIDKNKYETGILMNSGFLDGLSINRKYFLPGDTHGCONVSYGPGF  
 NLSFYNDKRYNDCGDYNNMGSGCYEYSASLSITPYKGNSTLAISSNTSYSTRYD  
 ANSEYPISTIKGRTRWQDLTAFTVVRMGDYNIMPTIGYNSQKQKQADGGLSLTL  
 TRYDGGKSLNAGSYNYSRQNTSYNDAFEGRLVSDTNVSYREL.SARVSGNRYTEGG  
 VSGRIINRGEDLNGTL.SYVKNRKSHTDTHSLTAGYSSFALTLDGTYWGSAGLTNL  
 SGCIYVAKSYNDESDLNTYKSSYGNYSGLSNDSLF.LPVALMQASLTIKENTKSN  
 IDVLAPTKNTFEMLP.GSVYDIDVANSVFTYVGRGVDRPLSGAYTINAQIVIDE  
 NGGEFSFESSENEKEFLDKDTIYSCSLDRSEARNQIVFGEVACNSTIKELLPEKLY  
 TNSRHLDLAYINDTE"  
 4121. 5212  
 /gene="csbD"





```

/gene="cotB"
/citation-[1]
/codon_start=1
/transl_table=1
/product="CotB"
/protein_id="CA87760.1"
/db_xref="GI:897726"
/db_xref="SPTREMBL:Q47116"
/transl_table="MKILLFVLEFNFAASNFVYDISKIDSGSGSETIKVESKSK
DVOYIKITKRVINGTEKEQVVDKNNMGGLIYTPARVLPAGASKSIRLIEIKKE
OEEVRYVESVKPGOODIEKNGRVTDSVNIYALIRISPEORLQDVIES
NNWIKNGTNRIGIKDYELCDTISINDKAKFVNNRLYPMDSVDTLGRKGFYAY
IDTKDDRENGSELINIKLP"
499.552
/gene="cotB"
/citation-[1]
553.1212
/product="CotB"
/gene="cotB"
/citation-[1]
1241.1245
/citation-[1]
1255.1767
/gene="cotA"
1255.1767
/gene="cotA"
/citation-[1]
/codon_start=1
/evidence-experimental
/transl_table=1
/product="CotA"
/protein_id="CA87761.1"
/db_xref="GI:897727"
/db_xref="SPTREMBL:Q47117"
/transl_table="MKLNKIGALVLSFTVSMGASAEKNITVASYPTIDLMQSD
GALPSAVNIAYLPEKERESARINTOVHTNNKTGQIDIKLNDVWVTNIDSPKTI
PLVSEFACTKLSTATSTITADQINGAAGVEVSAIKELVINAGSTQGTNIYAGNYGQ
LVSYLTNEP"
1253.1323
/gene="cotA"
/citation-[1]
/evidence-experimental
1324.1764
/gene="cotA"
/product="CotA"
/function="Major pilin subunit for CS2 pil1"
/citation-[1]
/evidence-experimental
1822.1824
/citation-[1]
1836.4436
/gene="cotC"
1836.4436
/gene="cotC"
/citation-[1]
/codon_start=1
/transl_table=1
/product="CotC"
/protein_id="CA87762.1"
/db_xref="GI:897728"
/db_xref="SPTREMBL:Q47118"
/transl_table="WRANKITVFLTLPLGLCFGTNGLESKKNIPDEFIDIMODEL
LEVNLYGRSLGVHRLTPTPTVKSVEIEKINVOKEDKEDLSLLQSSRNGM
SCNGDEKEYSCNTRIDTVANIVDEENNELIFGASFLVQADNDIYQKNISEK
AFISQITNESSESEGYSLKGVAGGLTENSYLFPMDAIYNSRKYTKNOSIN
IYRYDEMLKYIYLRMDRDLSSASGNFNENMLPLDIDGQICGTYOYIKRIEK
SISSEVYMLTRFSVEAFRNDELIGVYILNSGINDLDTSLPDGSDYILKIEPQDI
LVREKVPFNKGASFQDMQDVFAQGNIVNNDSYIEKQTKKGTGINAGIRPEVTR
NLSEFGAIIIDNKKYEGAVNMRSGFLVSGNFSLYGDAAGANTSYTDGFN
LSFYNDKSYVNCSHNYSAWNGCYEYSFSLVYSGMTTLGYNHNNNAVHYXD
TPEYFESKKYGVSKRMQLTSSSYKMDYHPIPIGYRSDQSWSGGGEISLFT
RVKENSAINAGSYNYVYKHNATHEAFIDGRTITNTPRGYSGLSINNNKNTAGVYI
GRVKNRFGDLNGLNVNKSITSGKMTSHMSANTNSFAITGDSYIMGDAAGLTKLISG

```

```

GVNVSRDSDSKELIKISGSSYGNVILGSDNDRSFLPVSAIMPNSMLTTEIOSDNKNIT
VOALSKNDFFILPGNVPEIDVYANVTYSIGRALDDGNPLSNNAHILDVHGRVLEDEG
GESEFSAKDFKSLFLKDKDIYSCDVKKYDIRSGVLEFGDILICHSIERLGRDLYVNN
PRVKOLAYK"
1836.1913
/gene="cotC"
/citation-[1]
1914.4433
/gene="cotC"
/product="CotC"
/citation-[1]
4441.4445
/citation-[1]
4451.5545
/gene="cotD"
4451.5545
/gene="cotD"
/note="t1g start codon"
/citation-[1]
/codon_start=1
/transl_table=1
/product="CotD"
/protein_id="CA87763.1"
/db_xref="GI:897729"
/db_xref="SPTREMBL:Q47119"
/transl_table="MKRVIFVLSMFLCSQVYGGQSMHNVNAGSINKTESIGPIDRSAA
ASYPAHYFHEHVAGYNNKDSLEFPMETLCSASDASKACPTENSKSSQGETNKL
IFTEKKSILARKTLNKGKRFLESYDRCHIVYVDAMNSHTKVCVSEFTGVDFTLYI
POGRIDGLLNGIWEATLELKRHYDYNHGTAVYNTVLTLDKGNIOVMPKPHSP
RIDNLNRPENKYSGSNVLEMCIXDGSHTSOSIEKRFQDSOTGNNEYNLKITGEP
LKLUPYKLSLLGGEHFPNNGEAFTINDISLFTINNRKISVSLPOISIPVLCWPAN
LTFMSELNLPAGEVSGILNVTFFPSSSL"
4451.4504
/gene="cotD"
/citation-[1]
4505.5542
/gene="cotD"
/product="CotD"
/citation-[1]

BASE COUNT      1904 a      874 c      1221 g      1799 t
ORIGIN

Query Match      27.5%; Score 298.8; DB 1; Length 5798;
Best Local Similarity 58.5%; Pred. No. 4e-54;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

127 GAATCTCCCCCAATATATATTAATTAATACCATTAACGATACAGTAAAGTCAT 186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4583 GCATCGTATCTCGTCATATATATTAATTAATGACATGTTGCTGTTACATTAAGATCAC 4642
187 ACTCTGATGATAGAGTACTTTTATATGTTGCTCTCTCAATATACACTTAATGAGCA 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4643 TCTCTTTTGAAGAGATGACGTTTATATGATGATCAACAGATGATCAATTAAGTGTCA 4702
247 TGTCCACACGATGACATCTAGCAGTTCAATCGGTCAAGCGTGAACAATATTAACATTA 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4703 TGTCCGACAGGAAACCTCCA-----AATCCTCTCAAGGGGAGACTAATATTAAGCTA 4756
307 CAATTAAGGAAAAAAGATTATATAAAGAGAGCAACAATTAAGGCTATTAACAA 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4757 ATATTACTGAAAAAAGTCTGGCCAGAAAAACATTAATTAAGATTAAGAGA 4816
367 TTATGTTCAAAAGTGAATGTC-----CCATCGGCTTAACACTTAATGACACT 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4817 TTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 CATTTAAGTGAATTAATAAAGCGGCTTCAGGTGCAAGTTTATATATATATTCCTGCT 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4877 ACTGTTAATGTGTAGTTCATTCACAAGAGAGATTAATTCACATTATATATCCACAA 4936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 GGCAGACAAAATTTTCCCTTTGGTGTATTCGGGATGCTACTGCAATTAAGAGTA 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4937 GGTGAATTTGATGGGCTTAACTGAGGTATATGGGAGGCAACATTAAGTTAGAGTGC 4996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

OY		538	AAAAGACGATATAGTGGACCTATGGAACTCAGTACCTTTAAATATCACATTAATAAATTACT	597
Dd		4997	AAAAAGCATTCACGCCTATATCATNGTNCCTAACAAAGTTAATATCACAGTTGATTGGACA	5056
OY		598	GATAAGGGAATATATTCAGATATATGGTTACTCCGTCAAAAGTGACGGTCGGCGATCTT	657
Dd		5057	GACAAAGAATATATTCAGGTCGTGGACACCAAAGTTTCATPAGCGATCCTAGAATTGATCGT	5116
OY		658	AACCTGGCTCACTGGTGGGGGACATATATTGGAAAGAAATTTGTTGATATGTGCTTT	717
Dd		5117	AATTACGCTCCGTAAGGTAATGTAAATATTCCTGTGTAGTAACTGCTGAGATGTGCTC	5176
OY		718	TATGATGATATATGATTAACAGCAGCTCTTTGAGATTAAGATTTCAAGATACAAATCGT	777
Dd		5177	TATGATGGCTATATAGTACACATATGTCACAAAGTATAGAAATGAGGTTTCAGATGACTCACAA	5236
OY		778	AAATCGATGAGGGAATTTTATCTTAAGGAAAAATTAATGATGACACCACAAAGAAATTCATAT	837
Dd		5237	ACAGGAATATATGATATATTAATCTTAATAAAAAAGTGAAGCCATTAATAAAAAATTCGCATAT	5296
OY		838	ACTTTGTCACTTCTCTTGGCGGGTAAAAAGTTTAATCTCCACAAATGGAAGCTCATTTAAT	897
Dd		5297	AAACTTTCCTCTCTTTAGAGAGCAGAGATTTTATCCAAATTAATGAGAGAGGCTTTTACT	5356
OY		898	ATTGCTGACCCAGCTTCTCTGGAACAAACATGGAATGAATTAACAGCGTGCACCATGGCA	957
Dd		5357	ATTATATATATCTCGTCTATGTGTTTATTAACCTGGAAATGATTAAGTCTGTATCTTTACCA	5416
OY		958	GAATTCAGTGTCCGGTGTGTGTGTGGCCCTGGACGTTTGCATTTGATGCAAAAGTGGAA	1017
Dd		5417	CAGATTGTATTTTCAGATACATATGCTGGCCAGCAAACTTGACATTTATGTCCAGACTTAAT	5476
OY		1018	AATCCGAGCGGTGGACAAATATATGGTAAATTTATGTTACTTTCACACCAAGTAGTCAA	10777
Dd		5477	AATCCAAGAGGGGTGATTAATCAGAAVACTTAACGTAACATTTATCTCTAGTAGTTCA	55386
OY		1078	ACACTCTTA 1085	
Dd		5537	AGTCTGTA 5544	
RESULT 10				
AX741415				
LOCUS	AX741415	5798 bp	DNA	Linear PAT 10-MAY-2003
DEFINITION	Sequence 5 from Patent WO03022306.			
ACCESSION	AX741415			
VERSION	AX741415.1	GI:30524208		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and Darsley,M.J.			
TITLE	Bacterial vaccine			
JOURNAL	Patent: WO 03022306-A 5 20-MAR-2003;			
FEATURES	Acambis Research Limited (GB)			
Source	Location/Qualifiers			
	1..5798			
	/organism="Escherichia coli"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:562"			
	499..1215			
	/gene="cotB"			
	499..1215			
	/gene="cotB"			
	/codon_start=1			
	/transl_table=1			
	/protein_id="CAD90909.1"			
	/db_xref="GI:30524209"			
	/translation="MKILLFYLIFENPAASANFMVYPISKIDIOGSGETIKVFSKR			

Query Match	Similarity	Score	DB	Length
Best Local	58.5%	Pred. No. 4e-54;		
Matches	566;	Conservative	0;	Mismatches 387; Indels 15; Gaps 2;
Db	4703	TGTCGACAGGAGAAACTCCA-----AATCCCTCAAGGGGAGACTAAATATTAACTTA	4756	
Db	247	TGTCACACAGTGAAGATTCCTAGACGTTCAATGCTTGTCTTCACAAATTAACCTTAATGACGA	306	
Db	4643	TCTCTTTTGGACAGATGACGTTTATNGATGATCATCAACAGATGACATTAAGTGCA	4702	
Db	187	ACTCTGATGATGAGATGACTTTTATNGTTTGTCTTCACAAATTAACCTTAATGACGA	246	
Db	4583	GCATCGATCTCTGCTCATATATATTTATGATCAACATGTTGCTGTTCACAAATTAAGATCAC	4642	
Db	127	GAATCTCCCCCAACATATATATTTAAATTAACCATATTAACGATATCAGTGAAGTCAT	186	
Db	4583	GCATCGATCTCTGCTCATATATATTTATGATCAACATGTTGCTGTTCACAAATTAAGATCAC	4642	
Db	187	ACTCTGATGATGAGATGACTTTTATNGTTTGTCTTCACAAATTAACCTTAATGACGA	246	
Db	4643	TCTCTTTTGGACAGATGACGTTTATNGATGATCATCAACAGATGACATTAAGTGCA	4702	
Db	247	TGTCACACAGTGAAGATTCCTAGACGTTCAATGCTTGTCTTCACAAATTAACCTTAATGACGA	306	
Db	4703	TGTCGACAGGAGAAACTCCA-----AATCCCTCAAGGGGAGACTAAATATTAACTTA	4756	

```

OY 307 CAATTACGAAAAAGAGTTTAAATAAGAGAGCTACAAATTAAAGCTATAACAA 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4757 AATTTTACTGAAAAAGAGCTGGCCAGAAAAAATTAACCTAAAGATATAAGACA 4816
OY 367 TTATGTCTCAAAAGTTTACTGC-----CCATCGCGCTTAACCTTAACGCT 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4817 TTTTATATGATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 4876
OY 418 CATTTAATCTGATATATATATATATATATATATATATATATATATATATATAT 477
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4877 ACTGTTAATCTGTAGCTTCATTCACAGAGAGATGATTTTACTTTATATATCCACA 4936
OY 478 GCGCACTTAAAAATTTGCTTTTGGTGTATCTGGAGTGTACTCTGATTAAGAGTA 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4937 GGTGAATATGATGGGCTTCTAATCTAGGATATATGAGGAGCAACTAGATGAGAGTC 4996
OY 538 AAAAAGCATATAGAGAGCTATGAGCTTACACTATATATATATATATATATATAT 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 4997 AAAAGGCACTTACGACTATATATATATATATATATATATATATATATATATAT 5056
OY 598 GATAAGGAAATATTCAGATATAGTTTACCTCAGTTCAAAAGTGAAGCTCGCTGATCTT 657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5057 GACAAAGGAAATATTCAGGCTGAGACACCAAGTTTCATAGGATCTGTATATGATCTG 5116
OY 658 AACTTGCCTCAACTGCTGGGGGACATATATATGAGAAATTCGTTGATATATGCTTT 717
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5117 AATTACGCTCTGAAAGGATATGATGATGATGATGATGATGATGATGATGATGATG 5176
OY 718 TATGATGATATATATATATATATATATATATATATATATATATATATATATAT 777
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5177 TATGATGCTATATATATATATATATATATATATATATATATATATATATATAT 5236
OY 778 AATCTGATGAGAAATTTTATCTAAGAAATTAATGATGACCAAAAGAAATTCATAT 837
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5237 ACAGAAATATATGATATATATATATATATATATATATATATATATATATATAT 5296
OY 838 ACTTGTCTACTCTCTGCGGGGTTAAAGTTTAACTCCAAATATGAAAGCTATATAT 897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5297 AAACCTTCTCTCTTTTGGAGAGAGAGATTTCACAAATATGAGAGAGCTTTTACT 5356
OY 898 ATTGCGAGCGAGCTTCTCTGGAACAAACCTGATATATATATATATATATATATAT 957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5357 ATTATGATGATCTCTCTCTATGTTTAACTGAAATCGATTTAGCTGTATATCTTACCA 5416
OY 958 GAATCACTGCTCGGCTGTGTGCTGCGCTGAGCTTTGCAATTTGAGTCAAAAGTGA 1017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5417 CAGATTAGATATCCAGTACTATGCTGCGCAGCAAACTGACATTTATGACAGCTAAT 5476
OY 1018 AATCCGAGGCTGACATATATATGTTATATATATATATATATATATATATATATAT 1077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5477 AATCCAGAGGCGGTGAGATTCAGGAATATCTTAACGTAACTTACTCTAGTAGTCA 5536
OY 1078 ACACCTCTA 1085
    ||||| |||||
DB 5537 AGTCTGTA 5544
    |||||

```

```

RESULT 11
AX083744/c 1141 bp DNA linear PAT 28-FEB-2001
LOCUS AX083744
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION AX083744
VERSION AX083744.1 GI:13185472
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Kunst, L. and Clemens, S.
AUTHORS Regulation of embryonic transcription in plants
TITLE Patent: WO 011061-A 22 15-FEB-2001;
JOURNAL UNIVERSITY OF BRITISH COLUMBIA (CA)

```

```

FEATURES
source
    Location/Qualifiers
    1..1141
    /organism="synthetic construct"
    /mol_type="genomic DNA"
    /db_xref="taxon:32630"
    1..1141
    /note="consensus sequence of A.t., L.a., and B.n. PAB1
    promoters"
BASE COUNT 123 a 32 c 42 g 112 t 832 others
ORIGIN
Query Match 5.8%; Score 63.4; DB 6; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.0018;
Matches 58; Conservative 247; Mismatches 239; Indels 3; Gaps 1;
OY 165 TACACATACAGTGAAGAGCTATGATGATGATGATGATGATGATGATGATGATGATG 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 633 TTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 574
OY 225 TCACATACACTTAATGAGAGATGTCACACAGTGAAGATGATGATGATGATGATGATG 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 573 KSAHSYWMYSNNAMRYRYSARNMS---SMAKRTTRNNMMMSGBVRMRAGTMMRMHM 517
OY 285 CGGTAAACAAATATATACATTTCAATTTACGAAAAAGAGTTTATATATATATATATAT 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 516 NNNTDTRYMMWRKARBTTTYVDSMCNAKSMWRGNRRNRRKMMWAAANNADGADMDHTYW 457
OY 345 ACAATTTAAAGGCTTAACAATTTATGTTCAAAAGTTTAACTGACATCGCGGCTTAC 404
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 456 MCNNTMMRRAMKMMNMWCRATCCNNNNRACVYHKKMKRTYKMKAAACNNNBK 397
OY 405 ACTTAATCAGCTCATTTTAACTGATATATATATATATATATATATATATATATATATAT 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 396 AMYKRVAMMYSRDNTNDMMWMTSDMEHMYTVDYTMRRANNNNNNNNNBCKTSSMW 337
OY 465 ATATTTCTCTGCTGCGGAGCAATATATATATATATATATATATATATATATATATAT 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 336 WMDHNTCTGCTGNTTWSAYBMAASMAAASBVTYNNWCVRMTYMKRTJNNNNNNKA 277
OY 525 GAAGTTAAGATTAATAAGAGATATAGTACAGCTATGAGCTTACATATATATATATAT 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 276 WYRTKTYAMCNRNRYVLTAVTWBKRNYKYCAUABWYBMGMHNBWBRABHRRSNM 217
OY 585 TATTAAATTAATCTGTAAGGAAATATATACATATGTTTACCTGATGTTCAAAAGTGA 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 216 WVKCRNRYNWSMHARBYRKBAYGCGNNNMWDMRAHHMCAATNNNNMMWYAYUHHM 157
OY 645 TCGCGTGAATCTTAATCTGCTGCAACGCTGCGGGGACATATATGGAAGAATCTGCT 704
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 156 HKKGAANTNNKTABRDDBAHVKTYYWRYDYWCAMCMWNAKAKVKTAMHMYTDRY 97
OY 705 TGATATG 711
    |||||
DB 96 VSANNTG 90
    |||||

```

```

RESULT 12
CEY50E8A 61864 bp DNA linear INV 21-MAY-2003
LOCUS CEY50E8A
DEFINITION Caenorhabditis elegans YAC y50E8A, complete sequence.
ACCESSION AL117200
VERSION AL117200.2 GI:6580305
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 none.
AUTHORS Genome sequence of the nematode C. elegans: a platform for
TITLE Investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613

```

PUBMED	9851916
REMARK	The C. elegans Sequencing Consortium
REFERENCE	2 (Pages 1 to 61864)
AUTHORS	Steward,C.A.
TITLE	Direct Submission
JOURNAL	Submitted (06-SEP-1999) Nematode Seq

[illegible]

Accession	Version	Keywords	Source	Organism
AF253047.1	GI:12082668			Candida albicans
AF253047				Candida albicans
AF253047				Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE				1 (bases 1 to 5230)
REFERENCE				Buurman,E.T., Jiang,W., McCoy,M., Averett,D.R., Thompson,C.M. and Wobbe,C.R.
TITLE				Validation of Cdc68p as a novel antifungal target
JOURNAL				Arch. Microbiol. 178 (6), 428-436 (2002)
MEDLINE				22306560
REFERENCE				12420162
REFERENCE				2 (bases 1 to 5230)
ADTHORS				Buurman,E.T., Jiang,W. and Wobbe,C.R.
ADTHORS				Direct Submission
JOURNAL				Submitted (06-APR-2000) Scriptogen Pharmaceuticals, 610 Lincoln Street, Waltham, MA 02451, USA
FEATURES				location/Qualifiers
source				1..5230
				/organism="Candida albicans"
				/mol_type="genomic DNA"
				/strain="1006"
				/db_xref="taxon:5476"
				complement(<1..>379)
				/gene="pr12"
				complement(<1..>379)
				/gene="pr12"
				complement(<1..>379)
				/product="DNA primase large subunit Pr12p"
				complement(<1..379)
				/gene="pr12"
				/codon_start=1
				/transl_table=12
				/product="DNA primase large subunit Pr12p"
				/protein_id="AAG48573.1"
				/db_xref="GI:12082669"
				/translation="MFQVQRRTAGRRNFEDSTATTYYTYYTNGEKDSPLYTPS
				ALVARSFDFDLPTQETLLEEFTEATIDRLKILIEISCLANSKTLKEIETSIKPLL
				LKYVLPNLTSGGCVNYEKERMKD"
				<834..>3992
				/gene="cdc68"
				<834..>3992
				/gene="cdc68"
				/product="large subunit Cp complex Cdc68p"
				834..3992
				/gene="cdc68"
				/note="Spt16p"
				/codon_start=1
				/transl_table=12
				/product="large subunit Cp complex Cdc68p"
				/protein_id="AAG48574.1"
				/db_xref="GI:12082670"
				/translation="MSEVNDIAGLFYFRLSIPQKLTANNIPQALITVGRSDNNYK
				KSYLVQWMLGEPYTHAIYITDQKCFITSEKSKHLKLTNOKPDLVETLRTQDV
				EHNKQLEIKLLETHKIDSKRGKIKDKTGKRTIDENQILNNNNNTIDHLSAVD
				LAVVSQALAVKDSSEFNNTKRIASMASVVMDFVNDMTIIVDEKRTINSQLTQIE
				DELENNWYDKTKRGKGLQSIKDFDEYLAECYSPILQGGDYDLKPSAVSTDPLI
				DGGLIYSSIGLRKYSYCNIAFTFLIDPTSEMETNYDFILQKRYIVDLKLGKVAS
				KYVQDITDIKKERPDNTVHNFTKNGCMLGMEPRDSFTILNATTPQKLTNGQISLT
				IGNNYISNDKNDKNDNTNOKKQYFALLITPTIKTDSLSLLTNYSKDRAISFS
				PNDNFEQKNNNNNNKRPGLSOTSNITGLKLESTETATLKLKRLHENTNADANSE
				KLQEQIITLHKRLOGLARFSKADTDADDFRPIFKYESTYRSQSPVNSVNDKLI
				HIDYKQTIILHKISGRVPHINSYKSGSONEGRDFYLLNFNSPAGNVAQKQEL
				PYEDSPNSFLRSTTIRSRORAVDYKAIDQLKDSYKREOKOMADVITQANLI
				ELKGSRYKLNWVEIRTPDTKKIGVLAIQHENGLRQSOPOSOSQNFENDRVDL
				FSNKLHFQPCQDELIVILHCLKNMIGKRTKTEQYQYRRAQMAPETGGRKK
				YRGDEDELOQOEDEERRKALLDKKEFGAEILADSSHGAVDIDIPRELGFQGVPR
				SVYLCVTRCLVOLIDPFLVYVLTLEIEIAHLERVOGKLNFDLVFEDEFKRPVH
				INILIPVELIDKSWLTDIPIISEGQMNINWQIKNTVYLDADYQFIIDGAGMLTGO
				GESDEEESDSFVRSVSDPDDEESDSVASEESDSYSGSDDDGGSGGDDGDE











## CDS

MIKELTSLQASIVTDCCKGELNEDGKKIKALMTDPGVKQMRANEFSLPDSAPYEF  
 DSIDRMTSPVYIPTQDILHTRKMTRGVHEINFEIGIKIRLVGVGQSRKRWLSC  
 EDDVAVVFCVALSEYDILLYEDNSTNRMLESLRVESDVCSNMFVNTPIILELNKSDL  
 FREKIKHVDLSETPPEYKGRDYERASNVYIKERFMQINKTBOKAIYSHITCATDTNNI  
 RVFEAVKDIIFTOCVMKAGYS"  
 join(22686, .22743,22820, .22926,23004, .23139,23232, .23558,  
 23639,27180)  
 /note="Genid exon scores (in order of location ranges):  
 -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ\_ID dd\_01857"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AA052224.1"  
 /db\_xref="GI:28829711"

## CDS

complement(join(27340, .29056,29330, .29490))  
 /note="Genid exon scores (in order of location ranges):  
 218.38, 14.76 - GSCJ\_ID dd\_01858"  
 /codon\_start=1  
 /product="similar to Rhizobium loti (Mesorhizobium loti).  
 Dehydrogenase, succinate semialdehyde dehydrogenase,"

Query Match 4.7%: Score 50.6; DB 3; Length 88549;  
 Best Local Similarity 49.8%; Pred. No. 0.56;  
 Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY	360	TAACAAATTATTTGTTCAAAAGGTGTAACGGCCATCCGCTTAACACTTAACCTCACTCA	419
DB	35734	TACCAATTAATTAATTAATGAATGATGATTCACCTTCATTCACACCTGATGATGAATC	35793
QY	420	TTTAACTGTATATAAAAGCGGCTTCAGTGCAGTTATTTATATATATCCGTGG	479
DB	35794	TTTCTATATTCAGAGAAATCAACAATATGTGAACCTAAATTTAAATATTTATATG	35853
QY	480	CGAAGTAAAAATTTGCTTTGTGTGATCTGGAGTGTCTACTCTGAAGTTAAGATAA	539
DB	35854	TGAATATATATATGTAATGATATTAATATCAAGATTAATATATATCAATATAAATTAAT	35913
QY	540	AAGAGATATAGTGAACCTATGACCTTACCTATAAATATCACTATTAATTAATG	599
DB	35914	ATTACCAATTAATGCAATCAATTAATTAATATATATATATATATATATTAATTAATAA	35973
QY	600	TAAGGAATATATTCAGA	616
DB	35974	TAATAATAATAATAATA	35990

Search completed: August 25, 2003, 18:03:00  
 Job time : 398 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:45:38 ; Search time 343 Seconds

(without alignments)  
8546.909 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086  
Sequence: 1 atgaataagattattat.....caagtagtcaaacacttag 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	AA170763	ETEC CS4 pilus csa
2	1086	100.0	7239	AA170780	ETEC CS4 pilus csa
3	298.8	27.5	5798	AA84848	CS2 gene cluster.
4	49.4	4.5	10279	ABU92277	Chemically treated
5	49.4	4.5	10279	ABU33591	Human immune syste
6	49.4	4.5	10279	AA22328	Chemically treated
7	48.2	4.4	9504	ABK28407	DNA transcription
8	47.2	4.3	5139	AA170139	Plasmodium falcipar

C	9	45.2	4.2	4176	23	ABU10866	Drosophila melanog
C	11	44.8	4.1	13133	24	ABK31231	Signal transductio
C	10	43.8	4.0	6971	24	ABN80061	Human chemically m
C	12	43.4	4.0	2036	23	ABU01847	Drosophila melanog
C	13	43.4	4.0	6080	23	ABU01846	Drosophila melanog
C	14	43.2	4.0	50000	24	ABU55643	AmpeV genome fragm
C	15	42.8	3.9	4985	24	ABU75107	Anopheles gambiae
C	16	42.8	3.9	6137	24	ABU34450	Chemically treated
C	17	42.6	3.9	6137	24	ABU34450	Human metastasis a
C	18	42.6	3.9	641	24	ABU56694	Human colon cancer
C	19	42.4	3.9	19734	24	ABU33933	Human immune syste
C	20	42.2	3.9	9095	24	ABK28448	DNA transcription
C	21	41.6	3.8	6209	22	AA528751	Genomic sequence #
C	22	41.6	3.8	15548	24	ABU34155	Human immune syste
C	23	41.6	3.8	2365589	24	ABA90521	Genomic sequence o
C	24	41.2	3.8	16228	24	ABU70459	Chemically treated
C	25	41.2	3.8	16228	24	AA561424	Human gene regulat
C	26	41.1	3.8	13420	24	ABU32917	Human immune syste
C	27	41.1	3.8	16842	22	AA546412	Tumour suppressor
C	28	41.1	3.8	16842	24	ABU70384	Chemically treated
C	29	41.1	3.8	16842	24	AA561336	Human gene regulat
C	30	41.1	3.8	16842	24	ABK31419	Signal transductio
C	31	40.8	3.8	6534	24	ABU32470	Human immune syste
C	32	40.6	3.7	73334	24	ABU92319	Chemically treated
C	33	40.6	3.7	73334	24	ABU34125	Human immune syste
C	34	40.2	3.7	34688	24	ABU67060	Human anglogenesis
C	35	40.2	3.7	6210	24	ABK28471	DNA transcription
C	36	39.8	3.7	9929	22	ABA17906	Human nervous syst
C	37	39.8	3.7	18878	22	AA137441	Human musculoskele
C	38	39.8	3.7	18878	25	ABK60429	CDNA encoding nove
C	39	39.8	3.7	47108	24	ABK31510	Signal transductio
C	40	39.8	3.7	74962	22	AAU15256	Human phosphatase
C	41	39.6	3.6	494	23	ABU10021	Human prostate exp
C	42	39.6	3.6	6175	24	AAU28373	Human chemically t
C	43	39.6	3.6	7676	24	ABU70409	Chemically treated
C	44	39.6	3.6	7676	24	ABU34598	Human metastasis a
C	45	39.6	3.6	15674	24	ABU70513	Chemically treated

#### ALIGNMENTS

RESULT 1	AA170763	AA170763 standard; DNA: 1086 BP.
ID	AA170763;	
AC	AA170763;	
XX	18-FEB-2002 (first entry)	
DT		
DE	ETEC CS4 pilus csaE gene coding region.	
XX		
XX	CS4 pilus; enterotoxigenic; ETEC; csa operon; csaE gene; fimbrial.	
KW	vacine; diarrhoea; antibacterial; antidiarrheic; ds.	
XX		
OS	Escherichia coli.	
XX		
FH	Key	Location/Qualifiers
FT	sig_peptide	1..69
FT	mat_peptide	70..1083
FT		/tag- a
FT		/tag- b
XX		
PN	W0200181582-A2.	
XX		
PD	01-NOV-2001.	
XX		
PF	20-APR-2001; 2001WO-US12914.	
XX		
PR	20-APR-2000; 2000US-198686P.	
XX		
PA	(UYMA-) UNIV MARYLAND BALTIMORE.	
XX		















[illegible]

CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAA70078 to AAA70287 and AAB1414 to AAB18352 represent nucleotide  
CC and protein sequences found in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 5139 BP: 2440 A; 434 C; 712 G; 1553 T; 0 other;

Sequence 5139 BP; 2440 A; 434 C; 712 G; 1553 T; 0 other;

Query Match	4.38;	Score 47.2;	DB 21;	Length 5139;
-------------	-------	-------------	--------	--------------

Matches 187; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

193 TATGATAGGATGACTTTTTTATGTTGTCTTCACATACACTTAATGGAGCATGTCCA 252

Db 997 TATTATAATATTTTAAATCCATTAAATGATAATTCAAATAATGTTGTTATTACA 1056

253 ACCAGTGAGAAATCCTAGCAGTTCATCGGTCAGCGGTGAACAATAATAACATTACAATT 312

Db 1057 ACTTGTGATAATAAGAAAGTTATCGAACACGACCAAGTGATATGATAAACAAATT TTTT 1116

313 ACGGAAAAAGAGCTTTAATAAAAGAGAGCTACAATTAAAGGCTATAACAATTATTG 372

Db 1117 GAAAGATGATGAATGAGAAAAAATATATTTAAAAATGAAAAATTTTAATGATGTTATT 1176

373 TTCAAAGTGTTAAGTGGCCATCCGGCCTAACACTTAAGTCACTCAGTCAATTTAACTGTAAT 432

Db 1177 AAAAAAATTCACATGCGAAAGAAAAATCTTAATTCAAATAGTACAATTATATG 1236

433 AAAAAGCGCGTTTCAAGGTGCAAGTTATATATTATTTCTGCGCGCAACGTAATAAT 102

hh	1	2	3	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00
hh	1	2	3	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00

**A03**

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

## RESULT 9

ID	ABL10866 standard; cDNA; 4176 BP.
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

AC ABL10866;

DT 26-MAR-2002 (first entry)

DE *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 27080.

Proscophila: developmental biology: insecticide:   
 kw   
 aa

KW pharmaceutical; gene; ss.  
XX

US *Drosophila melanogaster*.  
YY

PN WO2001/1042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX

XX

XX

DR WPI: 2001-656860/75.  
DR P-PSDB: ABB66763.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 27080; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABE01840-ABL16175) and the encoded proteins  
CC (ABE57737-ABE72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4176 BP; 1186 A; 862 C; 837 G; 1291 T; 0 other;  
XX  
Query Match 4.2%; Score 45.2; DB 23; Length 4176;  
Best Local Similarity 56.6%; Pred. No. 0.45;  
Matches 103; Conservative 0; Mismatches 78; Indels 1; Gaps 1;  
XX  
OY 689 TTGGAAGAAATCTGTGATATGCTTTATGATGATATAGTACTACAGCAGCTCTT 748  
DB 2504 TCGAAGGCAATATTTTAAATTTGTTCTGTATGGGTACATCAAAAAGAGCATTTT 2445  
OY 749 TGGAGATTAAGATTTCAGATTAACAATCTAAATCTGATGGCAAAATTTATCTTAAGGAAA 808  
DB 2444 AGCAATATCGCATCGATGCTATTAACCGTGCTCGTGAATAATATTCATATATGAAC 2385  
OY 809 TAAATGATGACACCAAGAAATTCATATACT-TTGCACTCTCTGGCGGTAAAGT 867  
DB 2384 TAAATTTAAGAACTAAGACATTTAAATATCTCTTAATTAATCTCTGCGCAGAAAT 2325  
OY 868 TT 869  
DB 2324 TT 2323  
XX  
RESULT 10  
ABK31231/C  
ABK31231 standard; DNA; 13133 BP.  
XX  
ABK31231;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Signal transduction associated gene modified complementary DNA #37.  
XX  
KM Human: signal transduction associated gene; cytosine methylation state;  
KM CpG island; signal transduction associated disease; solid tumour; cancer;  
KM antitumour; cytoskeletal; mutant; ds.  
OS Homo sapiens.  
OS Synthetic.  
XX  
XX WO200200926-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 29-JUN-2001; 2001WO-EP07472.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Plepenbrock C, Berlin K;  
PI

XX  
DR WPI: 2002-147896/19.  
XX  
PT Oligonucleotide for diagnosis and therapy of diseases associated with  
PT signal transduction e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with signal transduction -  
XX  
PS Claim 1; SEQ ID No 74; 24pp; English.  
XX  
CC The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically  
CC modified using a solution of bisulphite, hydrogen sulphite or  
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
CC for detecting the cytosine methylation state (CpG islands) of these  
CC genes, and a method for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with signal transduction.  
CC The genomic DNA can be obtained from cells or cellular components which  
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from  
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
CC histologic object slides, and all their possible combinations. The  
CC sequences of the invention are useful for the diagnosis and therapy of  
CC diseases associated with signal transduction e.g. solid tumours and  
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA  
CC sequences of different genes associated with signal transduction, or  
CC their complementary sequences.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 13133 BP; 3676 A; 350 C; 2780 G; 6326 T; 1 other;  
XX  
Query Match 4.1%; Score 44.8; DB 24; Length 13133;  
Best Local Similarity 44.9%; Pred. No. 0.79;  
Matches 169; Conservative 0; Mismatches 207; Indels 0; Gaps 0;  
XX  
OY 4 AATAAGATTTTATTTTATTTTACATGTTTCTCTTCAGTACCTTTTACATTGCTGTA 63  
DB 6047 AAAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5988  
OY 64 TCGGAGATTAATAATCCCGGAGATGAAGCATTAATATTTTGGCGCGGAGACAGG 123  
DB 5987 ACATTTAATCCCAAAACTAAATAATTCATTACACAAAATTAATTAATTAATTAATTA 5928  
OY 124 AACGAATCTCCCCCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 183  
DB 5927 ATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5868  
OY 184 CATACTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 243  
DB 5867 AATATTTTAAAAAATCCATTAATTTTAAAAATTTTAAAAATTTTAAAAATTAATCA 5808  
OY 244 GCATGTCACACAGAGAGATTCCTGACATTCATGCGGAGGAGGAGGAGGAGGAGGAGG 303  
DB 5807 ACAATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5748  
OY 304 TTACAATTTTACGAAAAAAGAGTTAATTAATTAATTAATTAATTAATTAATTAATTA 363  
DB 5747 ATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5688  
OY 364 CAATTAATTTCTCAAA 379  
DB 5687 CTTTACTTAATCAAAA 5672  
XX  
RESULT 11  
ABN80061/C  
ABN80061 standard; DNA; 6971 BP.  
XX  
ABN80061;  
XX  
DT 15-JUL-2002 (first entry)  
XX

DE Human chemically modified disease associated gene SEQ ID NO 78.

KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
antidiabetic; cytostatic; anticoagulant; da.

OS Homo sapiens.  
Synthetic.

PN WO200200927-A2.

PM 03-JAN-2002.

PP 02-JUL-2001; 2001WO-EP07536.

PR 30-JUN-2000; 2000DE-1032529.  
01-SEP-2000; 2000DE-1045826.

PX (EPIG-) EPIGENOMICS AG.

PA Olek A., Plepenbrock C., Berlin K;

PT WPI: 2002-130908/17.

XX

XX Novel nucleic acid useful for diagnosis and therapy of diseases  
associated with development genes such as diabetes, comprises a  
sequence of a segment of chemically pretreated DNA of genes associated  
with development

XX

XX Claim 1; SEQ ID NO 78; 27pp; English.

XX

XX The invention relates to a nucleic acid (I) comprising a sequence at  
least 18 bases in length of a segment of chemically pretreated DNA (II)  
of genes associated with development selected from 87 genes listed in  
the specification such as ACCRN, ADPW, or ARPD and comprising one of 350  
sequences (ABN79984-ABN80333), or their complements. The invention is  
useful for the diagnosis or therapy of diseases associated with  
development genes, in particular disease related to homeobox containing  
genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
associated with congenital heart disease, epilepsy, diseases related to  
histone deacetylation, Currarino syndrome, diseases related with the  
development of the brain and limb girdle muscular dystrophy and dwarfism.  
Oligomers specific to each of the genes are useful for detecting the  
methylation state of all CpG dinucleotides within the 350 sequences or  
(II) and their complementary sequences, as primer oligonucleotides for  
the amplification of the 350 sequences, (III) and/or their complements and  
as oligomer probes for detecting the cytosine methylation state and/or  
single nucleotide polymorphisms (SNPs).

XX Note: The sequence data for this patent did not form part of the printed  
specification but is based on sequence information supplied to Derwent by  
the European Patent Office.

XX

XX Sequence 6971 BP; 1786 A; 282 C; 1633 G; 3269 T; 1 other:

QY Query Match 4.0%; Score 43.8; DB 24; Length 6971;  
Best Local Similarity 47.0%; Pred. No. 1.1; Indels 0; Gaps 0  
Matches 135; Conservative 0; Mismatches 152;

Db 139 AAACATATAATTAAATTAACCATTTAGCAGTACAGTCAGTGAAAGCTACTGTGTATGAT 198  
||| |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
6783 AAAATATAAATCATATTTTATATACATTTTACCAAAAACCTGTATATGCATCTTTAAAAAAA 6722

QY 199 AGCATGACTTTTTATGTGGTCTCTTCACAAATACACGTTAATGAGCATGTCGAACCACT 258  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 6723 ATAAATTTATCTTTTATTAATAAACCACTTTTATATATTTATTTATTAACCTCCATTCCA 6666

QY 259 GAGATACCTACAGTTCATCGTCGCCGTGAAACAATATATTAACATTTACATTTGCGAA 318  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 6663 AATTTCCTCACCTCTCATATTTTAAACATATAACCAAAAARACATTAAATATAAATA 6600

QY 319 AAAAGAAGTTTATATATAAAGAGCTACAAATTTAAAGCCTATTAACATTTATGTTCAAA 378  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db	6603	TATAAAGCTTAAGCAAAAATAATTAATTAAATTAATCACTTTTATTAATAAAAAATNCAT	6544
QY	379	AGGTGTAACGCGCCATCCGGCTTAACACTTAAGCTACGCTATTTAA	425
Db	6543	AAATTTAATATTAACACCTTCACGTACACATATTAATTAAGCTAATTTAA	6497
RESULT 12			
ID	ABL01847/c		
XX	ABL01847 standard; cDNA; 2036 BP.		
XX	ABL01847;		
XX	26-MAR-2002 (first entry)		
XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 23.		
XX	Drosophila: developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; ss.		
XX	Drosophila melanogaster.		
XX	WO200171042-A2.		
XX	27-SEP-2001.		
XX	23-MAR-2001; 2001WO-US09231.		
XX	23-MAR-2000; 2000US-191637P.		
XX	11-JUL-2000; 2000US-0614150.		
XX	(PEKE ) PE CORP NY.		
XX	Venter JC, Adams M, Li PWD, Myers EW;		
XX	WPI; 2001-6556860/75.		
XX	P-PSDB; ABB57744.		
XX	New isolated nucleic acid detection reagent for detecting 1000 or more		
XX	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	interactions -		
XX	Claim 1; SEQ ID NO 23; 21bp + Sequence Listing; English.		
XX	The invention relates to an isolated nucleic acid detection reagent		
XX	capable of detecting 1000 or more genes from Drosophila. The invention is		
XX	useful in developmental biology and in elucidating cell signalling and		
XX	cell-cell interactions in higher eukaryotes for the development of		
XX	insecticides, therapeutics and pharmaceutical drugs. The invention		
XX	discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins		
XX	(ABB57737-ABB72072).		
XX	(ABB57737-ABB72072).		
XX	The sequence data for this patent did not form part of the printed		
XX	specification, but was obtained in electronic format directly from WIPO		
XX	at ftp://ipo.int/pub/published_pct_sequences.		
XX	Sequence 2036 BP; 703 A; 478 C; 385 G; 470 T; 0 other;		
QY	Query Match	4.0%; Score 43.4; DB 23; Length 2036;	
Db	Best Local Similarity	47.3%; Pred. No. 1;	
XX	Matches 131; Conservative 0; Mismatches 146; Indels 0; Gaps 0;		
QY	179	AAAGTCATCTCTGTATGATAGATGACTTTTATGTGTGCTTCACAAATACACTTA	238
Db	1980	ATATTTTTCGCCAACATCATCAATCAAAATGTTTCTTTGTTTATTTTACATCATGATAG	1921
QY	239	ATGAGAGATGCGCAACAGATGAGATGAGATCCTACACTTCATCGGTGACGGGTAAACCAATA	298
Db	1920	AGTGTTTTACCATCGAAATGAGAACAAATAATGTGTATTTTCGATTGCGGATCATTAATTA	1861
QY	299	TAACTATTCAATTTACGGAAAAAAGAGTTTAATTAATAAAGAGACTCAAAATTAAGGCT	358
Db	1860	AAATTCGATATATTTGCTTAAAAAACAACCTTGTCACAAAGATGCTCAGATCAGAGAAC	1801





OY 786 TGGGAATTATCTAGGAAAT 809  
| | | | | | | |  
Db 2369 TAATAATAATAATAATAATT 2346  
Search completed: August 25, 2003, 18:08:54  
Job time : 347 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:47:22 ; Search time 2242 Seconds

(without alignments)  
11772.819 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086

Sequence: 1 atgataagatttattat.....caagtagcaaaccttag 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues 45562784

Total number of hits satisfying chosen parameters:

Immun DB seq length: 0

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estlu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hnc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hnc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	59.4	5.5	1101	29	CNS0100X AL098379 Drosophila
C 2	55	5.1	1056	13	BX415058 BX415058
C 3	52	4.8	1200	13	BX415878 BX415878
C 4	51.6	4.8	1101	29	CNS00396 AL063921 Drosophila

Result No.	Score	Query Match	Length	ID	Description
C 5	51.2	4.7	1067	14	CD386564 AGENCOURT
C 6	51	4.7	1101	29	CNS00FEV AL071298 Drosophila
C 7	51	4.7	1225	29	CNS0161D AL106171 Drosophila
C 8	50.8	4.7	1124	13	BX436282 BX436282
C 9	50	4.6	1179	13	BX380916 BX380916
C 10	49	4.5	1043	29	CNS0145P AL107335 Drosophila
C 11	49	4.5	1061	13	BX437039 BX437039
C 12	48.6	4.5	1201	13	BX446296 BX446296
C 13	48.4	4.5	1201	13	BX355654 BX355654
C 14	48	4.4	1101	29	CNS0039G AL063921 Drosophila
C 15	47.6	4.4	931	14	CD389151 AGENCOURT
C 16	47.6	4.4	999	13	BX380865 BX380865
C 17	47.6	4.4	1101	29	CNS0145U AL103740 Drosophila
C 18	47.6	4.4	1148	13	BX360752 BX360752
C 19	47.4	4.4	925	18	BH152177 ENTOD23TF
C 20	47.4	4.4	1101	29	CNS000D1 AL065414 Drosophila
C 21	47.2	4.3	1101	29	CNS014XT AL104737 Drosophila
C 22	47.2	4.3	1204	29	CNS016E2 AL106628 Drosophila
C 23	47	4.3	838	29	CNS0096O AL053091 Drosophila
C 24	47	4.3	1101	29	CNS00FMC AL070972 Drosophila
C 25	46.8	4.3	1194	13	BX462185 BX462185
C 26	46.8	4.3	536	13	BQ400483 BQ400483
C 27	46.8	4.3	1101	29	CNS016HF AL106749 Drosophila
C 28	46.6	4.3	408	10	BE224677 BE224677
C 29	46.6	4.3	460	12	BI772835 BI772835
C 30	46.6	4.3	1201	13	AL532464 AL532464
C 31	46.6	4.3	1201	13	BX343343 BX343343
C 32	46.4	4.3	812	28	A2529785 ENTDA08TF
C 33	46.4	4.3	911	28	A2687106 ENT171TR
C 34	46.4	4.3	1092	29	CNS020K7 AL175696 Tetracton
C 35	46.4	4.3	1101	29	CNS000E8 AL063632 Drosophila
C 36	46.4	4.3	1101	29	CNS00EVL AL069706 Drosophila
C 37	46	4.2	427	13	BX403499 BX403499
C 38	46	4.2	1026	13	BX415239 BX415239
C 39	45.8	4.2	1201	9	AL536104 AL536104
C 40	45.6	4.2	852	29	CNS009AX AL053547 Drosophila
C 41	45.6	4.2	928	29	CNS00DKY AL071865 Drosophila
C 42	45.2	4.2	903	29	CNS017PV AL108349 Drosophila
C 43	45.2	4.2	994	13	BX414650 BX414650
C 44	45.2	4.2	1080	29	CNS00EPP AL069494 Drosophila
C 45	45	4.1	1001	29	CNS0155H AL105023 Drosophila

## ALIGNMENTS

RESULT 1  
LOCUS CNS0100X/C  
DEFINITION Drosophila melanogaster genome survey sequence sp6 end of BAC  
BACN03604 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL098379  
VERSION AL098379.1 GI:5609990  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector

FEATURES	Location/Qualifiers
source	1. .1101
	/organism="Drosophila melanogaster"
	/mol_type="genomic DNA"
	/db_xref="taxon:7227"
	/clone="BACN03604"
	/clone_11b="DrosBAC"
	/plasmid="PbeloBAC11"
	/note="end : SP6"
BASE COUNT	195 a 108 c 131 g 161 t 506 others
ORIGIN	
Query Match	5.5%; Score 59.4; DB 29; Length 1101;
Best Local Similarity	18.6%; Pred. No. 12;
Matches 116;	Conservative 245; Mismatches 257; Indels 7; Gaps 1;
2	TGAAATAGATTTTATTATTTTACATGGTTTCTCTCTGACATTTTTCATTTTGGTGC 61
1087	TAAATATAGATTTTATTATTTTACATGGTTTCTCTCTGACATTTTTCATTTTGGTGC 1028
62	TATCGGCAATAAATATTCGCGAGATGGAACATCAATATTTTTCGCGCGGATCA 121
1027	NATATATTCGCGAGATGGAACATCAATATTTTTCGCGCGGATCA 968
122	GGAAAGCAATTTTCCCGCAACATATATATTAATACCAATTTTACAGC-----ATAC 174
967	MCATCTTCT 908
175	AGTGAAGATCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
907	ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848
235	CTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294
847	CATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 788
295	AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
787	TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 728
355	GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 414
727	TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668
415	GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 474
667	TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 608
475	GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 534
607	AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 548
535	GTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 594
547	TTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
595	ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 619
487	CHHH 463
RESULT 2	
LOCUS	BK415058 1056 bp mRNA linear EST 15-MAY-2003
DEFINITION	BK415058 Homo sapiens THYMS Homo sapiens cDNA clone CS0CAP004Y619
ACCESSION	BK415058
VERSION	BK415058.1 GI:30767520
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

[illegible]





Db	858	DADWWTWDAADDWMAADDRWMDAMWKKMDAMWAGATARDRDMGDGAKRGKARRDRK	799
QY	439	CGCGCTTCAGGTGGCAAGTTATATATATATCTCGCTGGCGGAAGCTAAAAAATTGGCCT	498
Db	798	RADDKRDADRDADAATWTTTWTTCCTTTTBDTDDKMKWTDWTWRAAADTWTDWDDDDDDDRAG	739
QY	499	TTTGGTGATCTCGGATGCTACTCTGAAGTTAAAGATWAAAAGACATATAGTGAGACC	558
Db	738	TAGKRWRTWKRWMKRRDRTRMDNDADDTARDRRRRGDGDADAGKKTGKRRRRRDA	679
QY	559	TATGGAAGCTTACACTATAATATCACTATTTAAATTTAACTGATGAGGAAATATTCAGATA	618
Db	678	TWDRTDAAWMAAAWMTTDTDDDKDRBRKRGARRRRRTTAAAMDWMTWRAAMDWAM	619
QY	619	TGGTTACTCAGTCCAAAGTAGCCCTGCCGTCGATCTTAAGTTCGGTCCCACTGGTGGG	678
Db	618	DWKTRADRWMDWMAADTWTDARAKADDMKARWRKRRDRARARADRRTWTGKTTATW	559
QY	679	GGCAGATATATTTGGAAGAAATTCCTGTGATAGCTCTTTATAGATGATATAGTACTAAC	738
Db	558	TTWAKARAWMAWMAATATATATWTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWAA	499
QY	739	AGCAGCTCTTTGGAGATAGATTTTCAGGATAACAATCCTAAATCTGATGGGAAATTTTAT	798
Db	498	WTAAMAAAAAAMAAATTTTTTTTTTTTAAWMAAAWMTWMTWTTTWTWTWMAATTTT	439
QY	799	CTAAGCAAAATPAAATGATGACGCCAAAAGAAATTCATATACCTTTCGACTTTCG	852
Db	438	TWTTWTWMAATATATTTTWTWTWTWTAATAAAWMMWMTWATATKCCCCCCYCCC	385

[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1067)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Daniela S. Gerhardt, Ph.D.

National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: egapds-remail.nih.gov  
Tissue Procurement: Dr. Jamie Thompson, University of WI  
cDNA Library Preparation: Gina Zastrow-Hayes  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: NDKM74 row: e column: 02  
High quality sequence start: 8  
High quality sequence stop: 462.  
Location/Qualifiers  
I. 1067

```

/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/issue_type="Trophoblast"
/lab_host="DH10B Tona"
/clone_id="NH_MGC_173"
/note="Vector: pDONR201; Site 1: atp2; Site 2: atp1;
LIBR PRIMING - oligo dt; METHOD - full-length enriched;
LIBR PROVIDER - Bradfield"
BASE COUNT      533 a      85 c      75 g      33 t      41 others

```

ORIGIN	Query Match	4.7%;	Score 51.2;	DB 14;	Length 1067;
	Best Local Similarity	44.3%;	Pred. No. 2e+02;		
	Matches 155;	Conservative 0;	Mismatches 195;	Indels 0;	Gaps 0;
OY	290 AAACAAATATATACCTTACATTTCCGAAAAAGAGTTTAAATAAAGAGCTACAAA				349
Db	351 AA				410
OY	350 TTAAGGCTATAAACAATTTATGTCAAAAGTGTTAAGTCCATCCGGCTAACACTTA				409
Db	411 AA				470
OY	410 ACTCAGCTCATTTTACTGTATATAAAGCGGGCTTCAGTGCAGATTATATTATATYA				469
Db	471 ANNNNNNNNNNTTTTTTAAAAAAGGANNMAAAAAAAAAAANNAATTTTTTATTTT				530
OY	470 TTCCTGCTGGCGAAGCTAAAAAATTTGCCTTTGGGTGATGTCGGAGTCTACTCGAGT				529
Db	531 TTTTTTTTTTTNNAAAAAATTAATTTTTTTTTTTTATTTTAAAAAATATATTTTTTTTTT				590
OY	530 TTAAGATTAAGAGAGATATAGTGAAGCTATGGAAGTTACACTATAATATCACTATTA				589
Db	591 TTAAATATAAAAAATTAATTTAAANNNTAAAAATTAATTAANNAATTTATATTATAT				650
OY	590 AATTACTGATTAAGGAAATTTCTCAGATATGCTTACCTCAGTTCAAAAGT				639
Db	651 TAAATNTAAAAAAGAAAAATATATATTTTTTAAATATTAATTAATTAATTAATTAAT				700

RESULT_6	CNS00FE/c
LOCUS	CNS00FEV
DEFINITION	Drosophila melanogaster genome survey sequence TRJ3 end of BAC:
LOCUS	BACR3P8 of Rpct-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCSSION	AL071298
VERSION	AL071298.1 GI:4951138
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : :

COMMENT
<p>Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a>.</p>

FEATURES
<p>Location/Qualifiers</p> <p>1..1101</p>

```

FEATURES
  source      Location/Qualifiers
1. 1101
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /db_xref="taxon:7227"
   /clone="BACR32P18"
   /clone_id="RPC1.98"

```

with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCWSVSPORT 6 vector.



ACCESSION	fly), genomic survey sequence.
VERSION	AL103735
KEYWORDS	AL103735.1 GI:5615346
SOURCE	GSS.
ORGANISM	Drosophila melanogaster (fruit fly)
REFERENCE	Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
TITLE	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
JOURNAL	Ephydroidea; Drosophilidae; Drosophila.
COMMENT	1 (bases 1 to 1043) Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 J1006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pbeloBAC11.
FIGURES	Location/Qualifiers
SOURCE	1..1043

BASE COUNT	277 a	96 c	121 g	382 t	167 others
ORIGIN					

Query Match	4.5%	Score 49:	DB 29;	Length 1043;
Query Local Similarity	37.9%	Pred. No.	4.3e+02;	
Matches 136;	Conservative 48;	Mismatches 172;	Indels 3;	Gaps 1

[illegible]

RESULT	11
BX437039	
LOCUS	1061 bp mRNA linear EST_15-MAY-2003
DEFINITION	BX437039 Homo sapiens THYRDS Homo sapiens CDNA clone CSOCAF004TH04
ACCESSION	5-PRIME, mRNA sequence.
VERSION	BX437039
KEYWORDS	BX437039.1 GI:30789546
SOURCE	EST.
	Homo sapiens (human)

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 1061)
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91060 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr library was constructed by life technologies, a division of Invitrogen. this sequence belongs to sequence cluster 7009.f For more information about this cluster, see <a href="http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOCAP004DD02QP1&amp;cluster=7009.f">http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSOCAP004DD02QP1&amp;cluster=7009.f</a> . Contact : Feng Liang Email : <a href="mailto:liang@lifetech.com">liang@lifetech.com</a> URL : <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOCAP004DD02QP1. Location/Qualifiers 1. 1061
FEATURES	
SOURCE	

BASE COUNT	479 a	77 c	45 g	296 t	164 others
ORIGIN					

Query Match	4.5%	Score 49	DB 13	Length 1061
Best Local Similarity	42.2%	Pred. No. 4.3e+02		
Matches 139	Conservative 17	Mismatches 173	Indels 0	Gaps 0

[illegible]

RESULT 12	
BX446296	1201 bp mRNA linear EST 15-MAY-2003
LOCUS	BX446296 Homo sapiens NEUROBLASTOMA
DEFINITION	XCL08B001ZA03 3'-PRIME, mRNA sequence.
ACCESSION	BX446296
VERSION	BX446296.1 GI:30784407
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS L.L.W.B., Gruber.C., Jesse.J. and Polayes.D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: XCL0B0012A03BP1.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="XCL0B0012A03"  
/issue\_type="NEUROBLASTOMA"  
/clone\_lib="Homo sapiens NEUROBLASTOMA"  
/note="Vector: pCMVSPORT\_6; 1st strand cDNA was primed  
with a NotI-Oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 663 a 37 c 191 g 116 t 194 others  
ORIGIN  
Query Match 4.58; Score 48.6; DB 13; Length 1201;  
Best Local Similarity 35.08; Pred. No. 4.7e+02;  
Matches 129; Conservative 53; Mismatches 187; Indels 0; Gaps 0;  
QY 11 TTTTATTTATTTTACATGTTTCTCTCTCAGTACTTTTACATTTGCTGATCGGAG 70  
DB 152 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 211  
QY 71 ATAAATTTCCCGAGATGAAGCATTAATATTTTGGCCGCGTACAGAGAGCAT 130  
DB 212 WWWWAMMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMMWWMM 271  
QY 131 CTTCGCCCAACATATATATATTAATACCATTAATACGCTACAGTGAAGTCTACTC 190  
DB 272 ATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 331  
QY 191 TGTATGATAGATGACTTTTATGTTTCTCTCTCACAATACCTTAATGAGAGATGTC 250  
DB 332 AA 391  
QY 251 CAACCACTGAGATCGTACGATTCATCGGTGAGGAGTGAACAAATATTAATTAAT 310  
DB 392 AAAAAAAAAAR 451  
QY 311 TTACGGAAGAAAGATTATTAATAAGAGAGCTACAAATTAAGGCTATTAATTAAT 370  
DB 452 GRRRAAAAGGRRRAAAAGGRRRAAAAGGRRRAAAAGGRRRAAAAGGRRRAAA 511  
QY 371 TGTTCAAA 379  
DB 512 GRRRAAAA 520

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
AUTHORS L.L.W.B., Gruber.C., Jesse.J. and Polayes.D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CSODI002BE06BP1.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI002BE12"  
/issue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-Oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 493 a 49 c 96 g 240 t 323 others  
ORIGIN  
Query Match 4.58; Score 48.4; DB 13; Length 1201;  
Best Local Similarity 29.38; Pred. No. 5e+02;  
Matches 71; Conservative 72; Mismatches 99; Indels 0; Gaps 0;  
QY 226 CACATACCTTAATGAGATGTCACACAGTACGATTCATGCGTACGAG 285  
DB 1201 HMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMMAMMM 1142  
QY 286 GGTGAACAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 345  
DB 1141 MATHMTHTTMMMDMTTMMATTTTMMAMMAAAAMATTTTMMMAAAMMMMTTMM 1082  
QY 346 CAATTTAAGGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 405  
DB 1081 MMTTHTTMMTMMHMMTMMATTTTMMMAAAATTTTMMMAATTTTMMMAATTTM 1022  
QY 406 CTTAAGTACGCTATTTTAACTGTAATTAATAAGCGGCTCAGTCAAGTTATTTTA 465  
DB 1021 HMTTMTTMMATTTTMMATTTTMMATTTTMMATTTTMMATTTTMMATTTTMM 962  
QY 466 TA 467  
DB 961 AA 960

RESULT 14  
LOCUS CNS00396  
DEFINITION Drosophila melanogaster genome survey sequence TETS end of BAC #  
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit  
fly) genomic survey sequence.  
ACCESSION AL063921 GI:4941778  
VERSION AL063921  
KEYWORDS Drosophila melanogaster (fruit fly)  
SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1101)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :





**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:50:57 ; Search time 90 Seconds  
(without alignments)  
5326.023 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086  
Sequence: 1 atgaataagattttat.....caagtagtaaacactctag 1086

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	298.8	27.5	5798	2 US-08-483-101-1	Sequence 1, Appl1
2	39.8	3.7	74962	4 US-09-685-853A-3	Sequence 3, Appl1
3	39	3.6	10640	4 US-09-417-485D-5	Sequence 5, Appl1
4	38.8	3.6	342	4 US-09-601-198-115	Sequence 115, App
5	38.2	3.5	473	1 US-08-764-100-16	Sequence 16, Appl
6	38.2	3.5	4970	1 US-08-764-100-14	Sequence 14, Appl
7	38.2	3.5	4970	1 US-08-764-100-20	Sequence 20, Appl
8	38.2	3.5	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
9	37.4	3.4	533	4 US-09-495-050A-48	Sequence 48, Appl
10	36.6	3.4	2671	6 5168051-9	Patent No. 5168051
11	36.4	3.4	1020	4 US-09-463-962-1	Sequence 1, Appl1
12	36	3.4	540	4 US-09-404-879A-2	Sequence 2, Appl1
13	36	3.3	540	4 US-09-338-933-2	Sequence 2, Appl1
14	36	3.3	540	4 US-09-215-681-2	Sequence 2, Appl1
15	36	3.3	551	4 US-09-404-879A-137	Sequence 137, App
16	36	3.3	551	4 US-09-338-933-137	Sequence 137, App
17	36	3.3	551	4 US-09-215-681-137	Sequence 137, App
18	36	3.3	15016	4 US-09-601-198-60	Sequence 60, Appl
19	36	3.3	319608	4 US-09-539-333D-1	Sequence 1, Appl1
20	36	3.3	319608	4 US-09-679-409-1	Sequence 1, Appl1
21	35.6	3.3	1689	1 US-07-991-867B-41	Sequence 41, Appl
22	35.6	3.3	1689	2 US-08-544-332-41	Sequence 41, Appl
23	35.6	3.3	1689	4 US-09-370-861A-41	Sequence 41, Appl
24	35.6	3.3	1947	4 US-09-370-861A-74	Sequence 74, Appl
25	35.6	3.3	8457	1 US-07-991-867B-1	Sequence 1, Appl1
26	35.6	3.3	8457	2 US-08-544-332-1	Sequence 1, Appl1
27	35.6	3.3	8457	4 US-09-370-861A-1	Sequence 1, Appl1

28	35.6	3.3	19124	2 US-08-487-826B-13	Sequence 13, Appl
29	35.4	3.3	2826	4 US-08-624-655A-1	Sequence 1, Appl1
30	35.4	3.3	162450	4 US-09-345-882-1	Sequence 1, Appl1
31	35.4	3.3	169998	4 US-09-676-610B-24	Sequence 24, Appl
32	35.4	3.3	197496	4 US-09-877-177A-10	Sequence 10, Appl
33	35	3.2	1491	4 US-09-134-001C-866	Sequence 866, App
34	35	3.2	1170	4 US-08-961-527-172	Sequence 172, App
35	34.8	3.2	337	2 US-09-032-684-11	Sequence 11, Appl
36	34.8	3.2	612	4 US-09-107-532A-1950	Sequence 1950, Ap
37	34.6	3.2	4011	1 US-08-121-057-3	Sequence 3, Appl1
38	34.6	3.2	4011	2 US-08-509-187D-3	Sequence 3, Appl1
39	34.6	3.2	4011	2 US-09-121-396-3	Sequence 3, Appl1
40	34.6	3.2	4011	5 PCT-US93-09704A-3	Sequence 3, Appl1
41	34.6	3.2	4323	4 US-09-620-312D-134	Sequence 134, App
42	34.6	3.2	4838	3 US-08-852-629-15	Sequence 15, Appl
43	34.6	3.2	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
44	34.4	3.2	195	4 US-09-328-352-1980	Sequence 1980, Ap
45	34.4	3.2	801	4 US-09-134-001C-1507	Sequence 1507, Ap

ALIGNMENTS

RESULT 1  
US-08-483-101-1  
; Sequence 1, Application US/08483101  
; Patent No. 5932715  
; GENERAL INFORMATION:  
; APPLICANT: Scott, June R.  
; APPLICANT: Froehlich, Barbara  
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,101  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Farber, Donna M.  
; REGISTRATION NUMBER: 33878  
; REFERENCE/DOCKET NUMBER: 6-95  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5798 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 499..1215  
; FEATURE:  
; NAME/KEY: sig-peptide  
; LOCATION: 499..552

```

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 553..1212
FEATURE:
NAME/KEY: CDS
LOCATION: 1255..1767
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1255..1323
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1324..1764
FEATURE:
NAME/KEY: CDS
LOCATION: 1836..4436
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1836..1913
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1914..4433
FEATURE:
NAME/KEY: CDS
LOCATION: 4451..5545
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 4451..4504
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4505..5542
US-08-483-101-1

```

```

Query Match 27.5%; Score 298.8; DB 2; Length 5796;
Best Local Similarity 58.5%; Pred. No. 7.7e-70;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

```

```

OY 127 GAATGTCCTCCCAACATATATTAATTAACATATTAAGCATAGTAAAGTCT 186
DB 4583 GCATGTAATCTGCTCATATATATTTCTGTAACATGTTGGTTACATTAAGATC 4642
OY 187 ACTCTGTATGATAGATGACCTTTTATGTTGCTCTTCACAAATACCTTAATGAGCA 246
DB 4643 TCTCTTTTGAACAGATGACGTTTATGATATGATCATCAACAGATGATTAAGTGA 4702
OY 247 TGTCCACCATGAGAAATCTTACGATTCATGCTGACGGTGAAACAAATATAACATTA 306
DB 4703 TGTCCGACGAGGAAACTCCA-----AATCCTCTCAAGGGGAGACATAATTAAGCTA 4756
OY 307 CAATTTAGGAAAAAGAAAGTTTAATAAAGAGAGCTACAAATTAAGGCTATTAACAA 366
DB 4757 ATATTTACTGAAAAAGAAAGTGTGCCCAAAAAACATTAACCTTAAGGATATTAAGACA 4816
OY 367 TTATTTGTCAAAAGTGTAACTGC-----CCATCCGGCCCTTAACCTTAACCTGCT 417
DB 4817 TTTTATATGATATGATATGATATGATATGATATGATATGATATGATATGATATGAT 4876
OY 418 CATTTAACTGTAAATAAAAACGGGCTTCAGTGAAGTTTATATTAATTTCTGCT 477
DB 4877 ACTGTAAATGTGTAGTTTCAATTCACAAAGAGAGTATGATTTCACTTATATATCCACAA 4936
OY 478 GGGCAATAAAAAATTTGGCTTTTGGTATCTGGATGCTACCTGAAGTTAGAGA 537
DB 4937 GGTGAATATGATGGCTTCTTAATCTGAGATATATGGAGCAACCTAGAGTTAGAGTC 4996
OY 538 AAAAGCATATAGTACAGACCTATGAACTTACACTATTAATATCACTATTAATTAAT 597
DB 4997 AAAAGCATATAGTACATATATATGATATGATATGATATGATATGATATGATATG 5056
OY 598 GATTAAGGAAATATATGATATGATATGATATGATATGATATGATATGATATGATATG 657
DB 5057 GACAAAGGAAATATATGATATGATATGATATGATATGATATGATATGATATGATATG 5116
OY 658 AACTTGGCTCAACAGTGTGGGGGACATATATGGAAGAAATTTCTGTTGATATGCTT 717

```

```

DB 5117 AATTACGTCCTGGAAGTAATGTAATATTTCTGTTAGTAACTGCTGAGATGCTC 5176
OY 718 TATGATGATATAGTACTAAGACAGCTCTTTGGAGATTAAGATTCAGGATTAACATCT 777
DB 5177 TATGATGCTATAGTACATATGTAAGTATGAAATGAGATTCAGAGTACACAA 5236
OY 778 AATCTGATGGAAATTTATCTTAAGAAATTAATGATGACACCAAGAAATTCATAT 837
DB 5237 ACAGAAATTAATGATATATATCTTATTAATACTGAGACCATTAATAAATTTGCTAT 5296
OY 838 ACTTGTCACTTCTCTTGGCGGTTAAAGTTTAACTCCACAAATGAGCTATTAAT 897
DB 5297 AAACCTTCTCTTCTTTAGGAGAGCAAGATTTATCCAAATTAAGGAGAGCTTTTACT 5356
OY 898 ATTGCTGACGAGCTTCTCTGTAACAACTGATATGATATTAACATGCTGACATATGCA 957
DB 5357 ATTAATGATATCTTCTCTATTTTATTAATGGAATGCTATTAATGCTGATCTTACCA 5416
OY 958 GAATCAGTGTCCGCTGTGTTGTTGCTGACGCTTTCATTAATGATGACAAAGTGGA 1017
DB 5417 CAGATTAGATATTCAGTACTATATGCTGGCAGAACTTGACATTTATGTCAGAGCTAAT 5476
OY 1018 AATCCGAGCTGGACAAATATATGCTTAATTAATTTACTTTCACACCAAGTATCAA 1077
DB 5477 AATCCGAGAGCGGTAGTATTCAGAAATATTAACGTAATTAATTAATTAATTAAT 5536
OY 1078 ACACCTCA 1085
DB 5537 AGTCTGTA 5544

```

```

RESULT 2
US-09-685-853A-3/C
Sequence 3, Application us/09685853A
Patient No. 6479270
GENERAL INFORMATION:
APPLICANT: WEI, Ming et al.
TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
FILE REFERENCE: CLO00871
CURRENT APPLICATION NUMBER: US/09/685,853A
CURRENT FILING DATE: 2002-05-06
PRIOR APPLICATION NUMBER: 60/182,194
PRIOR FILING DATE: 2000-02-14
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO. 3
LENGTH: 74962
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(74962)
OTHER INFORMATION: n = A,T,C or G
US-09-685-853A-3

```

```

Query Match 3.7%; Score 39.8; DB 4; Length 74962;
Best Local Similarity 46.8%; Pred. No. 1.1;
Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
OY 105 TTTTGGCCCGCTGACAGAGAAAGATCTTCCCAACATATATATTAATTAACATAT 164
DB 43211 TTCTGACCTGATGACAGAGGAGACTCTGCTCAAAAAAAAAAAAAAAAAAAAA 43152
OY 165 TACACATACAGTGAAGTCACTCTATGATATGATGATGATTTTATATGTTGCTC 224
DB 43151 ATCACTGTTTGGGATATCTTTTAAATACCTATTTTAACTTAATGATGTTAA 43092
OY 225 TCACAATCACTTAATGAGATGTCACCAACAGTGAAGATCTGACATGATGCTGAG 284
DB 43091 AAAAAAAAACTACAAAGTAAGAGTCAAGCCCTCCATATATCCAAAGATTTCACTCTG 43032

```

1 APPLICATION NUMBER: US/08/214,066  
2 FILING DATE:  
3 APPLICATION NUMBER: US 08/032,233  
4 FILING DATE: 17-MAR-1993  
5 APPLICATION NUMBER: GB 9206016.9  
6 FILING DATE: 19-MAR-1992  
7 ATTORNEY/AGENT INFORMATION:  
8 NAME: No. 5773700c-1s, Allen E.

REGISTRATION NUMBER: 34,490  
 REFERENCE/DOCKET NUMBER: 137-1061  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 354-3592  
 TELEFAX: (415) 857-1125  
 INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 473 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-764-100-16

Query Match 3.5%; Score 38.2; DB 1; Length 473;  
 Best Local Similarity 43.8%; Pred. No. 0.5;  
 Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

420 TTTTACGTATATAAAGCGGCTTCAGTGCAGTTATATTATATATTCCTGCG 479
    || || || || || || || || || || || || || || || || || || ||
47 TTGAATGTTTAAATAAATAAAGCAAGATATAAATACTATATATATATATAGA 106
    || || || || || || || || || || || || || || || || || || ||
480 CGACTAAATAATTTGCTTTTGTGATCTGGATGCTACTGAACTTAAGATAA 539
    || || || || || || || || || || || || || || || || || || ||
107 AGTATATAATATATATATTTGTGTTTAAACAATCAAAACCAAAAAAG 166
    || || || || || || || || || || || || || || || || || || ||
540 AAGACGATATAGTGACCTATGAACTTACATATAATATCATTATTAATTA 599
    || || || || || || || || || || || || || || || || || || ||
167 AAAAAATAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATGAAA 226
    || || || || || || || || || || || || || || || || || || ||
600 TAAAGGAATATTCAGATATGTTACCTGACGTTCAAAAGTGCGTGGGATCT 659
    || || || || || || || || || || || || || || || || || || ||
227 AAGTTGAAAAAACCAAAAAATTTTGTGTAATAATAAGCTCGGCAAGTTGG 286
    || || || || || || || || || || || || || || || || || || ||
660 CTTCGCTCCAACTGGTGGGCGACATATATGGAATAATCTGTGATATGCT 719
    || || || || || || || || || || || || || || || || || || ||
287 TCTAAGACCTTTTATTTGTTTATACATTTTATTTGTTTGTGATTTTAT 346
    || || || || || || || || || || || || || || || || || || ||
720 TGATGATATATGATTAACAGAGCTCTTGGAGTAAAGTTTCAAGATACAA 779
    || || || || || || || || || || || || || || || || || || ||
347 TTAATTTATATATTTTATATAGTTGCTTATTAACCTTATTTAGAACAA 406
    || || || || || || || || || || || || || || || || || || ||
780 ATCTGATGGAATTTTAT 798
    || || || || || || || || || || || || || || || || || || ||
407 ATTGATTACATCATCTCT 425
    || || || || || || || || || || || || || || || || || || ||

```

SEQUENCE 6  
 08-764-100-14  
 Sequence 14, Application US/08764100  
 Patent No. 5773700  
 GENERAL INFORMATION:  
 APPLICANT: van Grinsven J., Martinus Q.  
 APPLICANT: De Haan, Petrus T.  
 APPLICANT: Gielen L., Johannes J.  
 APPLICANT: Peters, Dirk  
 APPLICANT: Goldbach, Robert W.  
 TITLE OF INVENTION: Improvements in or Relating to Organic  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sandoz Agro, Inc  
 STREET: 975 California Avenue  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/764,100

FILING DATE: 06-DEC-1996  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214,064  
 FILING DATE:  
 APPLICATION NUMBER: US 08/032,235  
 FILING DATE: 17-MAR-1993  
 APPLICATION NUMBER: GB 9206016.9  
 FILING DATE: 19-MAR-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5773700-15; Allen E.  
 REGISTRATION NUMBER: 34,490  
 REFERENCE/DOCKET NUMBER: 137-1061  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 354-3592  
 TELEFAX: (415) 857-1125  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4970 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-764-100-14

Query Match 3.5%; Score 38.2; DB 1; Length 4970;  
 Best Local Similarity 43.8%; Pred. No. 1.1;  
 Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

420 TTTTACGTATATAAAGCGGCTTCAGTGCAGTTATATTATATTCCTGCG 479
    || || || || || || || || || || || || || || || || || || ||
1044 TTGAATGTTTAAATAAATAAAGCAAGATATAAATACTATATATATATAGA 1103
    || || || || || || || || || || || || || || || || || || ||
480 CGACTAAATAATTTGCTTTTGTGATCTGGATGCTACTGAACTTAAGATAA 539
    || || || || || || || || || || || || || || || || || || ||
1104 AGTATATAATATATATATTTGTGTTTAAACAATCAAAACCAAAAAAG 1163
    || || || || || || || || || || || || || || || || || || ||
540 AAGACGATATAGTGACCTATGAACTTACATATAATATCATTATTAATTA 599
    || || || || || || || || || || || || || || || || || || ||
1164 AAAAAATAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATGAAA 1223
    || || || || || || || || || || || || || || || || || || ||
600 TAAAGGAATATTCAGATATGTTACCTGACGTTCAAAAGTGCGTGGGATCT 659
    || || || || || || || || || || || || || || || || || || ||
1224 AAGTTGAAAAAACCAAAAAATTTTGTGTAATAATAAGCTCGGCAAGTTGG 1283
    || || || || || || || || || || || || || || || || || || ||
660 CTTCGCTCCAACTGGTGGGCGACATATATGGAATAATCTGTGATATGCT 719
    || || || || || || || || || || || || || || || || || || ||
1284 TCTAAGACCTTTTATTTGTTTATACATTTTATTTGTTTGTGATTTTAT 1343
    || || || || || || || || || || || || || || || || || || ||
720 TGATGATATATGATTAACAGAGCTCTTGGAGTAAAGTTTCAAGATACAA 779
    || || || || || || || || || || || || || || || || || || ||
1344 TTAATTTATATTTTATATAGTTGCTTATTAACCTTATTTAGAACAAAT 1403
    || || || || || || || || || || || || || || || || || || ||
780 ATCTGATGGAATTTTAT 798
    || || || || || || || || || || || || || || || || || || ||
1404 ATTGATTACATCATCTCT 1422
    || || || || || || || || || || || || || || || || || || ||

```

SEQUENCE 7  
 US-08-764-100-20/c  
 Sequence 20, Application US/08764100  
 Patent No. 5773700  
 GENERAL INFORMATION:  
 APPLICANT: van Grinsven J., Martinus Q.  
 APPLICANT: De Haan, Petrus T.  
 APPLICANT: Gielen L., Johannes J.  
 APPLICANT: Peters, Dirk  
 APPLICANT: Goldbach, Robert W.  
 TITLE OF INVENTION: Improvements in or Relating to Organic  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sandoz Agro, Inc  
 STREET: 975 California Avenue

RESULT 8  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:

APPLICANT: Built et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methano  
Patent No. 6503729  
TITLE OF INVENTION: jannaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jannaschii  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (103398)..(103398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (234220)..(234220)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc\_feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g

```

NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741584)..(741584)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

```

```

LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match 3.5%; Score 38.2; DB 4; Length 1664976;
Best Local Similarity 51.5%; Pred. No. 8.2;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

QY 747 TTTGAGATTAAGATTTCAGATTAACAATCTTAATCTGATGGAATTTTATTAAGAA 806
DB 1486909 TGTGAATTTGATTAAGAAACAAATCCGTTTATGTTAAATTCCTATGAAAAACT 1486968
QY 807 AATAATATGACACCAAGAAATTCATTAATCTTGTCTTCTTGGCGGTTAAAG 866
DB 1486969 CTTAGCTATTAACCAAGATGATGATCTATTTGAATTAAGATTTGAGCGGCTT 1487028
QY 867 TTTACTCCAAACAAATGGAACGTCATTAATATTCGTCAGCGAGCTTCT 917
DB 1487029 TTTAATAAACAAGATATCTTTTATTTATGAAAAATACCAAAATCATCT 1487079

```

```

RESULT 9
US-09-495-050A-48
Sequence 48, Application US/09495050A
Patent No. 6492505
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATE
FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/09/495,050A
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/711,318
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 48
LENGTH: 533
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6492505 949387CT1
US-09-495-050A-48
Query Match 3.4%; Score 37.4; DB 4; Length 533;
Best Local Similarity 48.4%; Pred. No. 0.84;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

LOCUTION: (1)..(1020)  
US-09-463-962-1  
Query Match 3.4%; Score 36.4; DB 4; Length 1020;  
Best Local Similarity 52.7%; Pred. No. 1.9;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 140 AACATATATATTAATACCATTTACACATACAGTGAAGTCATCTCTGTATGATA 199  
D 132 AAGTATATGATTTCTTGAATAAATGCTAAGAGAGATTTAAATATCCTCACACAAAAA 191  
QY 200 GGATGACTTTTATGTTTCTTCTTCACAAATACCTTAATGGAGCATGCCAACCGTG 259  
D 192 TGATTAATATATGATTAATATATGTCATACGCTCAATTTAGCCATTCATATATGTA 251  
QY 260 AGATCCTACAGTTCATCGTCAGCGTGGAACAATATATACATTTACATTTACGGA 319  
D 252 TGCATGTTTCAAAACACATGTCGTACATATATATATATATTTTGTGAGTGAGA 311  
QY 320 AAGAGCTTAAT 354  
D 312 TAAATAGGTGTTAAAAAAGAAACAAAGAAAA 346  
RESULT 10  
5168051-9/c  
; Patent No. 5168051  
; APPLICANT: DERYNCK, RIK M.A.; GOEDDEL, DAVID V.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES  
; NUMBER OF SEQUENCES: 21  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/389,929  
; FILING DATE: 04-AUG-1989  
; SEQ ID NO: 9  
; LENGTH: 2671  
5168051-9  
Query Match 3.4%; Score 36.6; DB 6; Length 2671;  
Best Local Similarity 51.5%; Pred. No. 2.4;  
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 11 TTTTATTTATTTTACATTTTCTCTCAGTACTTTTACATTTGCTGATCGGAG 70  
D 2638 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2579  
QY 71 ATAAATTTCCCGAGATGAAGCATATATATTTTGGCCGCGTACAGAACGAT 130  
D 2578 ACCTTAGAATTAAGCCCTTTTGTGTTAGAAACAATCTTCCAGAAAGATGATCCAG 2519  
QY 131 CTTCCTCCCAACAT 173  
D 2518 ATTCCTCCCAACAT 2476  
RESULT 11  
US-09-463-962-1  
Sequence 1, Application US/09463962  
Patent No. 6413754  
GENERAL INFORMATION:  
; APPLICANT: FAYE, GERARD  
; APPLICANT: VALATY, JEAN-GABRIEL  
; APPLICANT: MANN, CARL  
; APPLICANT: THURET, JEAN-YVES  
; TITLE OF INVENTION: KINASE ACTIVATING DEPENDENT CYCLIN PROTEIN KINASES AND THEIR USES  
; FILE REFERENCE: 0846-0536-0XCT  
; CURRENT APPLICATION NUMBER: US/09/463,962  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: PCT/FR98/01788  
; PRIOR FILING DATE: 1998-08-11  
; PRIOR APPLICATION NUMBER: FR97/10287  
; PRIOR FILING DATE: 1997-08-12  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 1020  
; TYPE: DNA  
; ORGANISM: Candida albicans  
; FEATURE:  
; NAME/KEY: CDS

LOCUTION: (1)..(1020)  
US-09-463-962-1  
Query Match 3.4%; Score 36.4; DB 4; Length 1020;  
Best Local Similarity 52.7%; Pred. No. 1.9;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 679 GGACATATATATGAGAAATTTCTGATATGCTTTTATGATGATATATATATATAT 738  
D 790 GGTACACCAATTTTAACTGATTTTGAAGATGATATATTTTGTGATGATATATATAT 849  
QY 739 AGCAGCTTTTGGAGATTAAGATTTTACAGATTAACATCTTAAATCTGATGGAAATTTAT 798  
D 850 AACTGCTTTTAAAAATTTCAATTTTACAAATAATCTGAAAAAGATGGATATATAT 909  
QY 799 CTAGAAATATATGATGACCAAGAA 828  
D 910 TTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 939  
RESULT 12  
US-09-404-879A-2  
Sequence 2, Application US/09404879A  
Patent No. 648546  
GENERAL INFORMATION:  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.462C2  
; CURRENT APPLICATION NUMBER: US/09/404,879A  
; CURRENT FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 393  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 540  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-404-879A-2  
Query Match 3.3%; Score 36; DB 4; Length 540;  
Best Local Similarity 46.6%; Pred. No. 2;  
Matches 151; Conservative 0; Mismatches 170; Indels 3; Gaps 1;  
QY 318 AAAAGAGTTTAT 377  
D 66 AAGATGATTTAATATATGATTTTCACTTTTATCTGAGACAAAGTATCCATTA 125  
QY 378 AAGTTTACTGCCATCGGCTTACACTTAATCTGATTTTACTGATATATATAT 437  
D 126 TTAATGTCAGAAAGATGATTAACCTGCTTAAGAGCTTACGAAGCATATGGAGAG 185  
QY 438 CGGCGCTTCAGGTCGAAGTTTATATATATATATATATATATATATATATATATAT 497  
D 186 GTTGGCAGCAAGCAATTTTGAACATATATATATATATATATATATATATATATATAT 245  
QY 498 TTTTGG--TGATATCGGGATGCTACTGGAAGTTAAGGTAAGAGATATATCTGA 554  
D 246 TTTCGTGATGGAACTTATGAGCTTATTTGAAATGAGAGTTTACGAAAGCATGAGAC 305  
QY 555 GACCTATGAACTTACAT 614  
D 306 GCGAGACTGTGTATGATGCAATTAATGAAGCTTTATATGAAGTATATATATATATAT 365  
QY 615 GATATGTTTACCTGATCTCAAG 638  
D 366 AGCAGGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389  
RESULT 13  
US-09-338-933-2  
Sequence 2, Application US/09338933





Search completed: August 25, 2003, 18:48:16  
Job time : 97 secs

---

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 18:03:07 ; Search time 289 Seconds  
(without alignments)  
8448.117 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086  
Sequence: 1 atgaataagatttatat.....caagtagtcaacactctag 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Imm DB seq length: 0  
Imm DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	10	US-09-839-894-9
2	1086	100.0	7239	10	US-09-839-894-27
3	49.4	4.5	10279	12	US-10-311-455-1564
4	48.2	4.4	9504	12	US-10-240-453-281
5	43.2	4.0	499	13	US-10-027-632-287979
6	43.2	4.0	499	13	US-10-027-632-287980
7	42.8	3.9	4985	14	US-10-094-240-10
8	42.8	3.9	6137	12	US-10-240-485-3
9	42.4	3.9	19734	12	US-10-311-455-1906
10	42.2	3.9	9055	12	US-10-240-453-322
11	41.6	3.8	6209	9	US-09-764-860-1185
12	41.6	3.8	15548	12	US-10-074-095-1185
13	41.6	3.8	2241	13	US-10-311-455-2128
14	41.2	3.8	13420	12	US-10-108-605-116
15	41	3.8	6534	12	US-10-311-455-890
16	40.8	3.8	6534	12	US-10-311-455-443

17	40.6	3.7	7334	12	US-10-311-455-2098	Sequence 2098, Ap
18	40.2	3.7	598	14	US-10-102-524-1039	Sequence 1039, Ap
19	40	3.7	601	13	US-10-027-632-227464	Sequence 227464, Ap
20	40	3.7	601	13	US-10-027-632-227465	Sequence 227465, Ap
21	40	3.7	601	13	US-10-027-632-227466	Sequence 227466, Ap
22	40	3.7	6210	12	US-10-240-453-345	Sequence 345, Ap
23	39.8	3.7	985	13	US-10-027-632-151861	Sequence 151861, Ap
24	39.8	3.7	18878	14	US-09-764-877-3806	Sequence 3806, Ap
25	39.6	3.6	74962	12	US-10-274-974-3	Sequence 3, Ap
26	39.6	3.6	7676	12	US-10-240-485-151	Sequence 151, Ap
27	39.6	3.6	15674	12	US-10-311-455-335	Sequence 335, Ap
28	39.6	3.6	15674	12	US-10-240-485-29	Sequence 29, Ap
29	39.4	3.6	5268	12	US-10-311-455-850	Sequence 850, Ap
30	39.4	3.6	5572	12	US-10-311-455-1398	Sequence 1398, Ap
31	39.4	3.6	12507	12	US-10-311-455-271	Sequence 271, Ap
32	39.2	3.6	872	9	US-09-770-445-559	Sequence 559, Ap
33	39	3.6	5209	12	US-10-311-455-159	Sequence 159, Ap
34	39	3.6	5771	12	US-10-311-455-1923	Sequence 1923, Ap
35	39	3.6	6049	12	US-10-311-455-346	Sequence 346, Ap
36	39	3.6	10640	12	US-10-304-095-5	Sequence 5, Ap
37	39	3.6	19653	12	US-10-311-455-1307	Sequence 1307, Ap
38	38.8	3.6	720	11	US-09-791-279-54	Sequence 54, Ap
39	38.8	3.6	6375	12	US-10-311-455-1997	Sequence 1997, Ap
40	38.8	3.6	6668	12	US-10-311-455-1192	Sequence 1192, Ap
41	38.8	3.6	16724	12	US-10-311-455-1063	Sequence 1063, Ap
42	38.8	3.6	16724	12	US-10-240-485-89	Sequence 89, Ap
43	38.6	3.6	389	14	US-10-102-524-82	Sequence 82, Ap
44	38.6	3.6	4858	14	US-10-172-086-77	Sequence 77, Ap
45	38.6	3.6	9642	12	US-10-311-455-330	Sequence 330, Ap

ALIGNMENTS

RESULT 1  
US-09-839-894-9  
Sequence 9, Application US/09839894  
Patent No. US2002017686A1  
GENERAL INFORMATION:  
APPLICANT: Alboum, Zeev  
APPLICANT: Barry, Eileen M.  
APPLICANT: Levine, Myron M.  
TITLE OF INVENTION: Isolation and characterization of the  
FILE REFERENCE: USFMD 006A  
CURRENT APPLICATION NUMBER: US/09/839,894  
CURRENT FILING DATE: 2001-04-20  
PRIOR APPLICATION NUMBER: 60/198,626  
PRIOR FILING DATE: 2000-04-20  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 1086  
TYPE: DNA  
ORGANISM: E. coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1086)  
US-09-839-894-9

Query Match 100.0%; Score 1086; DB 10; Length 1086;  
Best Local Similarity 100.0%; Pred. No. 3.8e+248;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGAATAGATTATTTATTTTACATTTCTTCTTCACTACTTTTACATTTGCT 60
Db	1	ATGAATAGATTATTTATTTTACATTTCTTCTTCTTCACTACTTTTACATTTGCT 60
Qy	61	GTAATGCGACATTAATTCGCGAGATGAAGCATATTAATTTTGGCGCGTGAC 120
Db	61	GTAATGCGACATTAATTTCCGCGAGATGAAGCATATTAATTTTGGCGCGTGAC 120

```

QY 121 AGAAGCAATCTTCCCAACATATATATTAATTAACCATATTAACAGTACAGTGA 180
DB 121 AGAAGCAATCTTCCCAACATATATATTAATTAACCATATTAACAGTACAGTGA 180
QY 181 AGTCACTACTGTATGATAGATGACTTTTATATGTTGTCTTCACAAATCACTTAAT 240
DB 181 AGTCACTACTGTATGATAGATGACTTTTATATGTTGTCTTCACAAATCACTTAAT 240
QY 241 GGAGCATGTCCAACAGTAGAATCCTAGCAGTTACGCGTACAGCGGTGAACAATATA 300
DB 241 GGAGCATGTCCAACAGTAGAATCCTAGCAGTTACGCGTACAGCGGTGAACAATATA 300
QY 301 ACATTACATTTAGCGAAAAAAGAGTTTAATAAAAAGAGACTACAAATTAAGGCTAT 360
DB 301 ACATTACATTTAGCGAAAAAAGAGTTTAATAAAAAGAGACTACAAATTAAGGCTAT 360
QY 361 AAACATTTATTTGTCAAAAAGTGTACTGCTCCATCCGCGCTAACCTTAACCTACCTAT 420
DB 361 AAACATTTATTTGTCAAAAAGTGTACTGCTCCATCCGCGCTAACCTTAACCTACCTAT 420
QY 421 TTTAACTGTATTAATAAGCGGCTTCAGGTGCAAGTTTATATTAATTAATTCCTGCTGC 480
DB 421 TTTAACTGTATTAATAAGCGGCTTCAGGTGCAAGTTTATATTAATTAATTCCTGCTGC 480
QY 481 GAACATAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACCTGCAAGTTAAGTAAAGTAA 540
DB 481 GAACATAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACCTGCAAGTTAAGTAAAGTAA 540
QY 541 AGACGATATAGTAGACCTATGGAATTAACATTAATTAATTAATTAATTAATTAATTAAT 600
DB 541 AGACGATATAGTAGACCTATGGAATTAACATTAATTAATTAATTAATTAATTAATTAAT 600
QY 601 AAGGGAATATTCAGATATGTTACCTCAGTTCAAAAAGTGACGCTGCGGTGATCTTAAC 660
DB 601 AAGGGAATATTCAGATATGTTACCTCAGTTCAAAAAGTGACGCTGCGGTGATCTTAAC 660
QY 661 TTGCGTCCAACTGTGGGGGACATATATTTGGAAGAAATTCGTTGATATGCTTTAT 720
DB 661 TTGCGTCCAACTGTGGGGGACATATATTTGGAAGAAATTCGTTGATATGCTTTAT 720
QY 721 GATGATATATAGTAAACAGACGCTCTTGGAGATTAAGATTTACAGATTAACATCTTAA 780
DB 721 GATGATATATAGTAAACAGACGCTCTTGGAGATTAAGATTTACAGATTAACATCTTAA 780
QY 781 TGTGATGGAATTTTATCTAAGAAAAATTAATGATGACACCAAGAAATTCGATATCT 840
DB 781 TGTGATGGAATTTTATCTAAGAAAAATTAATGATGACACCAAGAAATTCGATATCT 840
QY 841 TTGTCACTCTCTGTGGCGGTAAGTTAACTCCACAAATGGAACGCTTAATATAT 900
DB 841 TTGTCACTCTCTGTGGCGGTAAGTTAACTCCACAAATGGAACGCTTAATATAT 900
QY 901 GCTGACGAGCTTCTCTGGAACAACAATGGAATTAAGATTTACAGCTGTCCACATGCCAGA 960
DB 901 GCTGACGAGCTTCTCTGGAACAACAATGGAATTAAGATTTACAGCTGTCCACATGCCAGA 960
QY 961 ATCAATGTTCCGGTGTGTGTGGCTGGAGCTTTGCAATTTGATGCAAAAGTGAAT 1020
DB 961 ATCAATGTTCCGGTGTGTGTGGCTGGAGCTTTGCAATTTGATGCAAAAGTGAAT 1020
QY 1021 CCCGAGGCTGACATATATGAGGTAAATTAATTTTACTTTTCACACCAAGTGAACA 1080
DB 1021 CCCGAGGCTGACATATATGAGGTAAATTAATTTTACTTTTCACACCAAGTGAACA 1080
QY 1081 CTCTAG 1086
DB 1081 CTCTAG 1086

```

RESULT 2  
 US-09-839-894-27  
 : Sequence 27, Application US/09839894  
 : Patent No. US2002017686A1

```

; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE OF INVENTION: CSA OPERON
; FILE REFERENCE: USFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 7239
; TYPE: DNA
; ORGANISM: E. COLI
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (283)...(999)
; NAME/KEY: CDS
; LOCATION: (1028)...(1531)
; NAME/KEY: CDS
; LOCATION: (1589)...(4192)
; NAME/KEY: CDS
; LOCATION: (4196)...(5281)
; NAME/KEY: CDS
; LOCATION: (5790)...(6119)
US-09-839-894-27

Query Match 100.08; Score 1086; DB 10; Length 7239;
Best Local Similarity 100.08; Pred. No. 8.7e-248;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATAGATTTTATTTATTTTACATTTGTTTCTGTCAGTACTTTTACATTTGCT 60
DB 4196 ATGAATAGATTTTATTTATTTTATTTTACATTTGTTTCTGTCAGTACTTTTACATTTGCT 4235

QY 61 GTATCGCAGATTAATAATTCGCGAGATGAAGCATTAATTTTGGCCGCGTGAC 120
DB 4256 GTATCGCAGATTAATAATTCGCGAGATGAAGCATTAATTTTGGCCGCGTGAC 4315

QY 121 AGAAGCAATCTTCCCAACATATATTAATTAATTAATTAATTAATTAATTAATTAAT 180
DB 4316 AGAAGCAATCTTCCCAACATATATTAATTAATTAATTAATTAATTAATTAATTAAT 4375

QY 181 AGTCACTACTGTATGATAGATGACTTTTATGTTGTCTTCACAAATCACTTAAT 240
DB 4376 AGTCACTACTGTATGATAGATGACTTTTATGTTGTCTTCACAAATCACTTAAT 4435

QY 241 GGAGCATGTCCAACAGTAGAATCCTAGCAGTTTCAAGTTCATGCTGACGGTGAACAATATA 300
DB 4436 GGAGCATGTCCAACAGTAGAATCCTAGCAGTTTCAAGTTCATGCTGACGGTGAACAATATA 4495

QY 301 ACATTACATTTAGCGAAAAAAGAGTTTAATAAAAAGAGCTACAAATTAAGGCTAT 360
DB 4496 ACATTACATTTAGCGAAAAAAGAGTTTAATAAAAAGAGCTACAAATTAAGGCTAT 4555

QY 361 AAACATTTATTTGTCAAAAAGTGTACTGCTCCATCCGCGCTAACCTTAACCTACCTAT 420
DB 4556 AAACATTTATTTGTCAAAAAGTGTACTGCTCCATCCGCGCTAACCTTAACCTACCTAT 4615

QY 421 TTTAACTGTATTAATAAGCGGCTTCAGGTGCAAGTTTATTAATTAATTAATTCCTGCTGC 480
DB 4616 TTTAACTGTATTAATAAGCGGCTTCAGGTGCAAGTTTATTAATTAATTAATTCCTGCTGC 4675

QY 481 GAACATAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACCTGCAAGTTAAGTAAAGTAA 540
DB 4676 GAACATAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACCTGCAAGTTAAGTAAAGTAA 4735

QY 541 AGACGATATAGTAGACCTATGGAATTAACATTAATTAATTAATTAATTAATTAATTAAT 600

```



QY 557 CCTATGAGACTTACACTATTAATATCATTATTAATTAAGTAAGGAAATATTGACA 616  
DB 8832 AATCACTAAAAAATCAATCTCTTAACAAATAATACCTACTAATCACTAAAAATA 8773  
QY 617 TAGGTACCCAGTCAAAAGTGA 641  
DB 8772 AATAATATCTCTCAATTTCTAAATAATA 8748

## RESULT 5

US-10-027-632-287979/c  
; Sequence 287979, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO: 287979  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287979

Query Match  
Best Local Similarity 4.0%; Score 43.2; DB 13; Length 499;  
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;

QY 669 AACTGGTGGGGACATATATTTGAGAAATTCGTGTATGCTTTATGATGAGATA 728  
DB 306 AACATTTGAGAGAAATTAATTTGACAAATTTGTTTAAATAGTTAATTTGTTATATA 247  
QY 729 TAGTACTAACACAGCTCTTTGGAGATAGATTTCAGGATACTTAATCTGATGCG 788  
DB 246 TAGTAGATGAGAAAGCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 187  
QY 789 GAAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCTGATGCTACT 848  
DB 186 TGAACCTCAATATAGTTTAAACCAAAACCAAAATTAATTAATTAATTAATTAATTA 127  
QY 849 TCTCTTGGGGGTAAGTTTAACCTCAACCAATGAGAGCTATTAATTAATTAATTAAT 908  
DB 126 TAGTGAACAGGTCTTTGATTACACCTCGAATCCACAGCATTTCAATATGCTTCAA 67  
QY 909 AGCTTCTCTG 918  
DB 66 AGTGTCTCTG 57

RESULT 6  
US-10-027-632-287980/c

; Sequence 287980, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO: 287980  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287980

; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO: 287980  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-287980

Query Match  
Best Local Similarity 4.0%; Score 43.2; DB 13; Length 499;  
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;

QY 669 AACTGGTGGGGACATATATTTGAGAAATTCGTGTATGCTTTATGATGAGATA 728  
DB 306 AACATTTGAGAGAAATTAATTTGACAAATTTGTTTAAATAGTTAATTTGTTATATA 247  
QY 729 TAGTACTAACACAGCTCTTTGGAGATAGATTTCAGGATACTTAATCTGATGCG 788  
DB 246 TAGTAGATGAGAAAGCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 187  
QY 789 GAAATTTTATCTAAGAAATTAATGATGACACCAAGAAATTCATATCTGATGCTACT 848  
DB 186 TGAACCTCAATATAGTTTAAACCAAAACCAAAATTAATTAATTAATTAATTAATTA 127  
QY 849 TCTCTTGGGGGTAAGTTTAACCTCAACCAATGAGAGCTATTAATTAATTAATTAAT 908  
DB 126 TAGTGAACAGGTCTTTGATTACACCTCGAATCCACAGCATTTCAATATGCTTCAA 67  
QY 909 AGCTTCTCTG 918  
DB 66 AGTGTCTCTG 57

RESULT 7  
US-10-094-240-10/c

; Sequence 10, Application US/10094240  
; Publication No. US20030082637A1  
; GENERAL INFORMATION:  
; APPLICANT: ZWIBEL, LAURENCE J.  
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF  
; FILE REFERENCE: N8289  
; CURRENT APPLICATION NUMBER: US/10/094,240  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 10/056,405  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: 60/264,649  
; PRIOR FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 10  
; LENGTH: 4985  
; TYPE: DNA  
; ORGANISM: Anopheles gambiae  
US-10-094-240-10

Query Match  
3.9%; Score 42.8; DB 14; Length 4985;



DB 3012 AACCTTAAACAAATTTAACTTAAATATAC--AAAAACAAAAAATAAATATAT 2955  
OY 343 CTACAAATTAAGGCTATTAACATTTATGTTCAAAAGTGTAACTGCCATCGGCCA 402  
DB 2954 CTAAACATTAATTAATTAACAAACAAACAAACAAACAAACAAACAAACAAAC 2895  
OY 403 ACCTTAACTGCGCTATTTTAACTGTAATTAACAAACGCGCTGCAAGTTATAT 462  
DB 2894 ACCTTAACTGCGCTATTTTAACTGTAATTAACAAACGCGCTGCAAGTTATAT 2835  
OY 463 TATATATCTGCTGCGCACTAAATAATTTGCTTTGGTATCGATCGATGCTACT 522  
DB 2834 TTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2775  
OY 523 CTGAAGTTAAGATAAAGACATATAGTGAACCTATGACCTATACATTAATATC 582  
DB 2774 CAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2715  
OY 583 ACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 612  
DB 2714 ACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2685

## RESULT 10

US-10-240-453-322/c  
; Sequence 322, Application US/10240453  
; Publication No. US20030148326A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
; TITLE OF INVENTION: Transcription  
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
; FILE REFERENCE: 5013.1009  
; CURRENT APPLICATION NUMBER: US/10/240,453  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03973  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 350  
; SEQ ID NO 322  
; LENGTH: 9095  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-322

Query Match 3.9%; Score 42.2; DB 12; Length 9095;  
Best Local Similarity 45.8%; Pred. No. 11;  
Matches 146; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

OY 522 TCTGAAGTTAAGATAAAGACATATAGTGAACCTATGACCTATACATTAATAT 581  
DB 6264 TCTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6205  
OY 582 CACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 641  
DB 6204 AACTCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6145  
OY 642 CGCTGCGCTGATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 701  
DB 6144 ATAAACCTTATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6085

OY 702 TGTGATATGCTTTTATGATGATATAGTAAACAGACAGCTTTGGAGATAGAT 761  
DB 6084 TATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6025  
OY 762 TCAGATTAACATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 821  
DB 6024 ATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5965  
OY 822 CAAGAAATTCATATATCT 840  
DB 5964 CATCACTCTTATATAT 5946

## RESULT 11

US-09-764-860-1185/c  
; Sequence 1185, Application US/09764860  
; Patent No. US20020094953A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008  
; CURRENT APPLICATION NUMBER: US/09/764,860  
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1198  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1185  
; LENGTH: 6209  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-860-1185

Query Match 3.8%; Score 41.6; DB 9; Length 6209;  
Best Local Similarity 51.0%; Pred. No. 13;  
Matches 98; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

OY 2 TGAATTAAGATTTTATTTTATTTTACATTTGTTTCTCTGAGTACTTTTACATTTGCTG 61  
DB 2688 TGTCTTAAGATTTTCTCTTTTATTTTACCTTTCTGTAATGAGAGTGTGAATTTTGTCA 2630  
OY 62 TATGGGAGATTAATTTCCCGGAGATGAAGATACATTAATTTTGGCCGCTGACA 121  
DB 2629 AACATATTTCTGAAATTTATTAATTAATTAATTTCTTTTATTAATTAATTAATTA 2570  
OY 122 GGAAGAAATCTTCCCAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 181  
DB 2569 CTGACTGATTTTCTTTTAACATTTTACTGTTTATTTTAAAAAATAGAGCAAG 2510  
OY 182 GTCATACTCTGT 193  
DB 2509 GTCTCAGCTTGT 2498

## RESULT 12

US-10-074-095-1185/c  
; Sequence 1185, Application US/10074095  
; Publication No. US2003007704A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008C1  
; CURRENT APPLICATION NUMBER: US/10/074,095  
; CURRENT FILING DATE: 2002-02-14  
; PRIOR APPLICATION NUMBER: 09/764,860  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487



;;  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214

```

? PRIOR FILING DATE: 2001-07-02
? PRIOR APPLICATION NUMBER: DE 10032529.7
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: DE 10043826.1
? PRIOR FILING DATE: 2000-09-01
? NUMBER OF SEQ ID NOS: 2424
? SEQ ID NO 2128
? LENGTH: 15548
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2128

Query Match          3.8%  Score 41.6:  DB 12:  Length 15548;
Best Local Similarity 48.3%  Pred. No.19;
Matches 116:  Conservative 0;  Mismatches 124;  Indels 0;  Gaps 0;

```

[illegible]

```

; ORGANISM: Drosophila melanogaster
us-10-108-605-116

Query Match      3.8%; Score 41.2; DB 13; Length 2241;
Best Local Similarity 47.6%; Pred. No. 10;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY    202 ATGACCTTTTAACTGTGGTCCTCCACAATACACTAATGAGCATGTCACCAACAGTGAG 261
      ||| |||| |
Db     2099 ATGTTTCTTTTGTGTTTTTTTTTTTACATCATGATTAAGTGTTTTTACCATTGCCAAATGAG 2040

QY    262 AATCTTAGCAGTTCATCGGTACGGGTGAACAATAATTAACATTACAAATTTACGAAAAA 321
      ||| |||| |
Db     2039 AACAAAATCTCTAATTTGCATTGCGCATCATATAATTTAAATTCGATTAATATGCTAAAAA 1980

```



**This Page Blank (uspto)**